Social Markers of Sexual Orientation and Gender in Speech and Appearance:

A Combination of Producer- and Perceiver-Centered Approaches

vom Promotionsausschuss des Fachbereichs 8: Psychologie

der Universität Koblenz-Landau

zur Verleihung des akademischen Grades

Doktor der Philosophie (Dr. phil.) genehmigte

Dissertation

Dipl.-Psych. Sven Kachel

Tag der Abgabe der Dissertation: 11.07.2017

Tag der Disputation: 07.11.2017

Vorsitzende des Promotionsausschusses: Prof. Dr. Melanie C. Steffens (Universität Koblenz-Landau)

Vorsitzende der Promotionskommission: Jun.-Prof. Dr. Meike Kroneisen (Universität Koblenz-Landau)

Berichterstatter*innen: Prof. Dr. Melanie C. Steffens (Universität Koblenz-Landau), Prof. Dr. Adrian P. Simpson (Friedrich-Schiller-Universität Jena), Prof. Dr. Gerhard Reese (Universität Koblenz-Landau)

Für den Vater.
Für den Kater.
Nicht für die Katz‘.

Table of Contents

List of Figures........III

Abbreviations........IV

Danksagung........V

Zusammenfassung........VIII

Abstract........XI

1. Introduction........1
   - 1.1. Sexual Orientation Stereotypes........3
   - 1.2. Gender-Role Self-Concept........7
   - 1.3. Acoustic Correlates of Gender........9
     - 1.3.1. Gender Differences and a Crash Course in Speech Production........12
     - 1.3.2. Masculinity/Femininity and a Classification of Sociophonetic Studies........15
   - 1.4. Expression of Sexual Orientation........18
   - 1.5. Perception of Sexual Orientation........22
   - 1.6. Aims of the Present Research........24
2. The Present Research........26
   - 2.1. Manuscript 1........27
   - 2.2. Manuscript 2........29
   - 2.3. Manuscript 3........31
   - 2.4. Manuscript 4........33
   - 2.5. Manuscript 5........34

3. General Discussion. 37

3.1. Locating the Results. 37

3.1.1. Stereotype (In)Accuracy. 37

3.1.2. Expression and Perception of Sexual Orientation Model. 39

3.1.3. Separating Biological and Social Effects of Gender on Speech. 41

3.1.4. Novel Methods in the Sociophonetics of Sexual Orientation. 42

3.2. Strength of the Present Research. 44

3.3. Limitations of the Present Research and Future Directions. 45

3.4. Conclusion. 48

4. References. 49

5. Appendix. 59

5.1. List of Manuscripts. 59

5.2. Manuscript 1. 60

5.3. Manuscript 1 – Related Appendix. 80

5.4. Manuscript 2. 82

5.5. Manuscript 2 – Related Appendix. 100

5.6. Manuscript 3. 110

5.7. Manuscript 3 – Related Appendix. 159

5.8. Manuscript 4. 176

5.9. Manuscript 5. 206

5.10. Manuscript 5 – Related Appendix. 278

Curriculum Vitae. 282

Eidesstattliche Erklärung. 285

List of Figures

Figure 1. Source-Filter Model of Speech Production 11

Figure 2. Gender Differences in Acoustic Vowel Space 13

Figure 3. Relations of the Single Research Aims and Manuscripts. 26

Abbreviations

cm	Centimeter
CoG	Center of gravity
EPSOM	Expression and Perception of Sexual Orientation Model
f0	Fundamental frequency
F1	First formant frequency
F2	Second formant frequency
Hz	Hertz
TMF	Traditional Masculinity-Femininity Scale

Danksagung

Die ersten Worte sind entscheidend. Sowohl deren Inhalt, als auch deren Form. Denn die Form bestimmt den Inhalt mit. So verwende ich lieber „Danksagung“ statt „Acknowledgments“, um eine sprachliche Nähe zu denen herzustellen, die mir nahe sind. Trotz des englischen Titels, der die Tür aufmachen soll für eine internationale Öffentlichkeit, soll ein kleiner Raum gewahrt bleiben, in dem das Private verhandelt werden kann. Die Menschen, die unmittelbar und mittelbar zum Beginnen, Werden und Beenden dieser Dissertation beitragen haben, mögen sich sprachlich ein- und nicht ausgeschlossen fühlen.

So wie meine Eltern. Ich danke meiner Mutti, die den Grundstein für diese Arbeit gelegt hat. Für ihre jahrelange und Krisen überdauernde, wohlwollende Unterstützung, dass sie mir (un)bewusst Modell für sorgfältiges, präzises und strukturiertes Arbeiten war und dass sie es wie wohl keine Zweite versteht, progressive und traditionelle Geschlechterrollenelemente in einer Person zu vereinen. Ihr allein hätte es bedurft. Und doch kam der entscheidende Impuls zur Beschäftigung mit dem Thema meiner Dissertation von meinem Vater. Kurz nach meinem Coming-Out sagte er: „Dann war die Sprecherziehung ja für umsonst“. Nicht schmeichelhaft, aber authentisch. Mit seinem Kommentar spielte mein Vater darauf an, dass ich mich nach meinem Coming-Out stärker an schwulen Sprechstereotypen orientieren und damit die Ergebnisse der Sprecherziehung konterkarieren würde, die ich während meines Studiums genießen durfte. Diesen wirkmächtigen Stereotypen nachzugehen und sie auf ihren wahren Gehalt hin abzuklopfen, war initiales Ziel der Arbeit. Dafür danke ich ihm.

Die primäre Wichtigkeit von Frauen als Rollenmodellen, spiegelt sich auch in meiner wissenschaftlichen Sozialisation wider. Ich danke Prof. Melanie C. Steffens nicht nur für ihre exzellente Betreuung und allumfassende Expertise zur Konzeption, Durchführung, Auswertung und Dokumentation von Studien, sondern auch dafür, dass ich erfahren konnte, wie versiert und scheinbar mühelos sich verschiedene Projekte und Aufgaben gleichzeitig jonglieren lassen, dafür, dass sie mir großes Vertrauen für meine erbetene Selbständigkeit im Arbeiten entgegenbrachte, und dass sie mit mir auf Augenhöhe kommuniziert hat. Prof. Adrian P. Simpson danke ich für sein überragendes Engagement als Dozent*in, welches die praxisorientierteste und mitreißendste Vorlesung zeitigte, die ich in meiner Studienzeit erleben durfte. Angesteckt durch dieses Feuer, fing auch ich an zu brennen für einen Bereich, der sich beharrlich und hingebungsvoll den kleinsten und meist vergessenen Elementen einer Sprache widmet: Den Lauten. Ohne seine kontinuierlichen Hinweise zu deren Messung und physiologischen Grundlagen, wäre diese Arbeit weit weniger phonetisch fundiert geblieben. Dafür und für seine Betreuung danke ich ihm. Prof. Gerhard Reese und Jun.-Prof. Meike Kroneisen danke ich, dass sie sich kurzerhand bereiterklärten, die Promotionskommission zu komplettieren.

Trotz meines eremitisch anmutenden Stils als Arbeiter*in im stillen Kämmerlein, war für mich der Kontakt zu meinen Kolleg*innen und Freund*innen aus dem wissenschaftlichen Umfeld essenziell. Ich möchte zuallererst Dr. Helene Kreysa danken, die sich trotz Elternzeit und ohne mit der Wimper zu zucken, der Korrektur meiner Dissertation angenommen hat. Ich danke Dr. Romi Zäske und Dr. Verena G. Skuk für eine Einführung in neue Methoden der Stimmforschung und dadurch entstandene Synergieeffekte. Alle drei gaben mir Orientierung, wenn ich sie brauchte, und fanden Zeit, um auch persönliche Belange zu besprechen. Dr. Claudia Menzel danke ich für ihre allzeit offenen Ohren und zusammen verbrachte Mittage und Abende. Dr. Franziska Ehrke, Claudia Niedlich, Theresa Prokein, Sabine Preuß und Sylvia Hohwiller danke ich für fachlichen Input, aber noch mehr für privaten Austausch mit Schokoriegel und Kakao sowie für das gemeinsame Lösen von Boulderproblemen.

Obwohl die Dissertation einen großen Platz in meinem Leben eingenommen hat, war es doch nie monothematisch. Dafür haben großartige Menschen gesorgt, die mich auf verschiedene Arten intellektuell stimulierten. Menschen bei denen ich mich aufgehoben und angenommen fühlte. Die ein sicheres Netz gesponnen haben, in welches ich mich freimütig fallen lassen durfte, wenn ich haderte. Menschen, für die es in Ordnung war, egal für welche Tür ich mich entschied. Mit denen echte Begegnung möglich war und ist. Ohne Maske. Katharina Korb, Dirk Hertrampf, Suzanne Friedrich, Julia Schubert, Sabine Reichelt, Anja Heinemann, Heike-Jana Stödtler, Christiane und Martin Beims, Kerstin Hartung, Anne Fekete, Christin Haucke und einige mehr: Ich danke euch für euren steten Rückhalt und eure Freundschaft. Johanna Schmidt, Dr. Julia Femel und Alexander Melzer gingen auf dem Weg leider verloren. Ich freute mich, fänden wir uns wieder. Die intensiven Gespräche und Auseinandersetzungen mit euch, haben mich nachhaltig inspiriert, verortet und reifen lassen.

Marko Junghanß – der Namen haben wir viele füreinander. Ich danke dir für die Kontinuität, die du in mein Leben gebracht hast, für deinen Witz und deinen Charme, deine Schlagfertigkeit und Selbstsicherheit, für Berührung und deine Schulter. Ich bin sehr glücklich, dass du es bist, mit dem ich durch’s Leben schreite. Mau, mau!

Zusammenfassung

Homonegative Diskriminierung, wie die Aberkennung von Führungsqualitäten oder die Verweigerung einer besseren Bezahlung, betrifft nicht nur Lesben und Schwule, sondern Menschen, die als lesbisch oder schwul wahrgenommen werden (Fasoli et al., 2017). Folglich wird angenommen, dass insbesondere heterosexuelle Personen Opfer homonegativer Diskriminierung werden (Plöderl, 2014). Zwar ist die Wahrnehmung der sexuellen Orientierung stereotypengetrieben (z.B. Cox et al., 2015), jedoch fehlt es bislang an Wissen darüber, wie korrekt diese Stereotype, insbesondere bezogen auf die Sprechweise, sind. Trotz diverser soziophonetischer und sozialpsychologischer Forschung im Zusammenhang mit sexueller Orientierung und Geschlecht, mangelt es bislang an einem umfassenden Verständnis, wie die sexuelle Orientierung ausgedrückt und wahrgenommen wird.

Diese Lücken möchte die vorliegende Arbeit schließen. Hauptziele der vorliegenden Arbeit sind die Überprüfung der Korrektheit von Sprechstereotypen im Kontext der sexuellen Orientierung (*Hauptforschungsziel 1*) und die Entwicklung eines Modells, wie die sexuelle Orientierung interpersonell konstruiert wird (*Hauptforschungsziel 2*). Um allgemein zu einem Mehr an Wissen über die Sprechweise und die äußere Erscheinung als soziale Marker der sexuellen Orientierung und des Geschlechts beizutragen, werden biologische und soziale Effekte des Geschlechts auf die Sprechweise getrennt untersucht (*Nebenforschungsziel 1*), neue methodische und technologische Verfahren in die soziophonetische Erforschung der sexuellen Orientierung eingebracht (z.B. nuancierte psychologische Analyse, Nasometrie, Voice Morphing; *Nebenforschungsziel 2*) sowie kommunikator*innen- und rezipient*innenzentrierte Ansätze kombiniert.

Die vorliegende Arbeit besteht aus insgesamt fünf Manuskripten. Diesen ist gemein, dass sie sozialpsychologische und soziophonetische Perspektiven integrativ behandeln, die soziale Identität in den Blick nehmen und primär die Sprechweise anstelle der fazialen Erscheinung in den Mittelpunkt stellen. Zudem wurden vornehmlich deutsche respektive deutschsprachige Versuchspersonen untersucht.

In Manuskript 1 wird die *Traditional Masculinity/Femininity-Scale* (TMF) als reliables und valides Instrument zur Erfassung des Geschlechtsrollen-Selbstkonzeptes etabliert. Diese

Neuentwicklung ist notwendig, da bislang existierende Skalen die selbstzugeschriebene Maskulinität/Femininität nur noch unzureichend abbilden (z.B. Abele, 2003; Evers & Sieverding, 2014). Die TMF repräsentiert durch nur sechs Items ein eindimensionales Maskulinitäts-/Femininitätskonstrukt als soziales Geschlecht. Dadurch ermöglicht sie die angestrebte Trennung von Effekten des biologischen und sozialen Geschlechts auf die Sprechweise, steht als neue Messmethode für die soziophonetische und allgemeine Geschlechterforschung zur Verfügung und bereitet daher die beiden Hauptforschungsziele vor.

In den Manuskripten 2, 3 und 4 wird die Korrektheit von Sprechstereotypen bezüglich ihres stereotypen Inhalts und der unterstellten Homogenität von Mitgliedern der gleichen Gruppe adressiert. Dies erfolgt durch den Einsatz unterschiedlicher methodischer Zugänge. Zum einen werden, einer konventionellen Vorgehensweise folgend, relevante akustische Parameter von Lesben und heterosexuellen Frauen (Manuskript 2) respektive von schwulen und heterosexuellen Männern (Manuskript 3) gruppenspezifisch gemittelt und miteinander verglichen. Zum anderen werden je fünf Repräsentant*innen von Lesben, Schwulen, heterosexuellen Frauen und Männern so ausgewählt, dass sich die Gruppen maximal voneinander unterscheiden (Manuskript 4). Aus den ausgewählten Einzelstimmen werden durch die Anwendung von Voice-Morphing-Verfahren prototypische, natürlich klingende Durchschnittsstimmen hergestellt (Kawahara et al., 2008), die Hörer*innen vorgespielt werden. Lesben und heterosexuelle Frauen (Manuskript 2) unterscheiden sich in keinem, schwule und heterosexuelle Männer nur in einem der analysierten akustischen Parameter (Manuskript 3): Heterosexuelle Männer produzierten Vokale mit einer geringeren akustischen Entfernung vom Munddach als schwule Männer. Eine nasometrische Untersuchung zeigte keine Unterschiede zwischen den Männergruppen. Demgegenüber erbrachte eine nuancierte psychologische Analyse verschiedentlich Evidenz für eine akustische Heterogenität innerhalb der Gruppen. Insbesondere die Exklusivität der sexuellen Orientierung wie auch das Geschlechtsrollen-Selbstkonzept wurden akustisch indexikalisiert. Dies legt nahe, dass Sprechstereotype inkorrekt sind. Jedoch wurde gezeigt, dass in den Durchschnittsstimmen die Informationen über die sexuelle Orientierung für Rezipient*innen wahrnehmbar enthalten sind (Manuskript 4). Dadurch lassen sich Sprechstereotype als Übertreibungen kleiner Körnchen von Wahrheit begreifen.

In Manuskript 5 wird die bisherige Literatur zur interpersonellen Konstruktion der sexuellen Orientierung zu einem Modell verdichtet: Dem *Expression and Perception of Sexual Orientation Model* (EPSOM). Dieses Modell beschreibt mit einer indirekten Route, wie die Information der sexuellen Orientierung von der Kommunikator*in hin zur Rezipient*in durch drei mediierende Komponenten übertragen wird. Diese indirekte Route wurde empirisch nur für Frauen als Kommunikator*innen nachgewiesen, nicht aber für Männer. Frauen weisen eine bestimmte *tatsächliche sexuelle Orientierung* (z.B. heterosexuelle Selbstidentifikation) auf, die bestimmte *psychologische Merkmale* wahrscheinlicher machen (z.B. geschlechtskonformere Selbsteinschätzung), welche sich in *impliziten Signalen* widerspiegeln (z.B. geschlechtskonformere Sprechweise). Diese Signale werden von Rezipient*innen als *Eindrücke, die mit der sexuellen Orientierung in Verbindung stehen,* dekodiert (z.B. Eindruck der Geschlechtskonformität), woraus letztlich die *wahrgenommene sexuelle Orientierung* abgeleitet wird (z.B. Wahrnehmung von Heterosexualität). EPSOM trägt zum einen der Heterogenität innerhalb der sexuellen Orientierungsgruppen Rechnung (durch die Aufnahme psychologischer Merkmale als Komponente) und deutet zum anderen darauf hin, wie die Information über die sexuelle Orientierung von der Kommunikator*in zur Rezipient*in verloren gehen kann. Dadurch ist EPSOM in der Lage, eine Erklärung dafür anzubieten, warum die sexuelle Orientierung zwar überzufällig gut aber nicht perfekt erkannt werden kann.

Insgesamt stellt die vorliegende Arbeit bedeutsame Impulse zur Weiterentwicklung der Erforschung von sozialen Markern der sexuellen Orientierung und des Geschlechts bereit. Sie schlägt ein (Denk)Modell vor, wie sexuelle Orientierung ausgedrückt und wahrgenommen wird, sie zeigt die Fruchtbarkeit der Verquickung sozialpsychologischer und soziophonetischer Forschungsansätze auf und verdeutlicht den Wert der Anwendung neuer Methoden und Technologien. Darüber hinaus weist die vorliegende Arbeit auch praktische Implikationen auf. Sprechstereotype im Kontext der sexuellen Orientierung können als inkorrekt zurückgewiesen werden – so sprechen deutschsprachige heterosexuelle Männer nicht mehr oder weniger nasal als schwule Männer. Gemäß EPSOM kann daraus, dass eine Person als lesbisch oder schwul wahrgenommen wird, nicht geschlussfolgert werden, dass sie auch lesbisch oder schwul ist. Damit trägt die vorliegende Arbeit potentiell zum Abbau von Stereotypen und zur Reduktion von Diskriminierung bei.

Abstract

Homonegative discrimination such as the denial of leadership qualities and higher salaries concern not only lesbians and gay men but also individuals who were perceived as lesbian or gay (Fasoli et al., 2017). Hence, it is assumed that especially straight people become victims of homonegative discrimination (Plöderl, 2014). The perception of sexual orientation is indeed stereotype-driven (e.g., Cox et al., 2015) but there is a lack of knowledge on how accurate stereotypes are – particularly those referring to speech. Despite a variety of sociophonetic and social psychological research related to sexual orientation and gender, an encompassing understanding is missing on how sexual orientation is expressed and perceived.

The present thesis aims to fill these gaps. The two major aims of the present work are the examination of the accuracy of speech stereotypes in the context of sexual orientation (Major Research Aim 1) and the development of a model on how sexual orientation is interpersonally construed (Major Research Aim 2). In order to provide more knowledge on speech and appearance as social markers of sexual orientation and gender, biological and social effects of gender and speech are investigated separately (Minor Research Aim 1), new methodological and technological procedures in sociophonetic research on sexual orientation are introduced (e.g., fine-grained psychological analysis, nasometry, voice morphing; Minor Research Aim 2), and producer- and perceiver-centered approaches are combined.

Overall, the present thesis comprises five manuscripts with the following aspects in common: They integratively deal with social psychological and sociophonetic perspectives, share a social identity approach, and primarily center speech instead of facial appearance. Moreover, mostly German and German native speaking participants, respectively, have been investigated.

Manuscript 1 establishes the *Traditional Masculinity/Femininity-Scale* (TMF) as a reliable and valid instrument for assessing gender-role self-concept. The invention was necessary because existing scales insufficiently represented the self-ascribed masculinity/femininity yet (e.g., Abele, 2003; Evers & Sieverding, 2014). The TMF indicates

a one-dimensional masculinity/femininity construct as social gender by using six items only. Thereby, the TMF enables the intended separation of biological and social effects of gender on speech, provides a new measurement for sociophonetic and gender research in general, and prepares the two major research aims.

Manuscripts 2, 3, and 4 address the (in)accuracy of speech stereotypes regarding stereotypic content and suggested within-group homogeneity. This is carried out by the application of different methodological approaches. On the one hand, following conventional modus operandi, relevant acoustic parameters of lesbians and straight women (Manuscript 2) and accordingly gay and straight men (Manuscript 3) were averaged for each group and compared between groups. On the other hand, five representatives of each lesbian/gay and straight women and men were selected in a manner that groups were maximally different from each other (Manuscript 4). Based on the selected single voices, prototypical and naturally sounding voice averages were created by applying voice morphing (Kawahara et al., 2008) and presented to listeners. Lesbians and straight women (Manuscript 2) differed in none, gay and straight men in one of the analyzed acoustic parameters only (Manuscript 3): Straight men produced vowels with a lower acoustic distance to the hard palate than gay men. By utilizing nasometry, both male groups showed no nasalance differences. In contrast, a fine-grained psychological analysis yielded various evidence for acoustic within-group heterogeneity. In particular, the exclusivity of sexual orientation and gender-role self-concept have been acoustically indexicalized which suggests that speech stereotypes are inaccurate. However, voice averages do carry perceivable sexual orientation information (Manuscript 4). Hence, speech stereotypes can be considered as exaggerations of tiny kernels of truth.

In Manuscript 5, previous literature on the interpersonal construction of sexual orientation is integrated in a model: The *Expression and Perception of Sexual Orientation Model* (EPSOM). This model postulates an indirect route and describes how sexual orientation information is transmitted from producer to perceiver by proposing three mediating components. The indirect route has been empirically confirmed for female but not male producers. Women show a certain *actual sexual orientation* (e.g., straight self-identification) that makes particular *psychological characteristics* more likely (e.g., gender conforming self-assessment) that are reflected in *implicit signals* (e.g., gender conforming speech). These signals are decoded by perceivers as *impressions related to sexual orientation* (e.g., impression of gender conformity) whereof finally *perceived sexual orientation* is derived (e.g., perception of straightness). EPSOM accounts for heterogeneity within sexual orientation groups by including *psychological characteristics* as a mediating component and indicates how sexual orientation information get lost on the way from producer to perceiver. Thereby, the EPSOM is able to offer an explanation why sexual orientation can be perceived with above-chance but far-away-from-perfect accuracy.

Overall, the present thesis provides meaningful impulses for enhancements of research on social markers of sexual orientation and gender. This thesis offers a model on how sexual orientation is expressed and perceived, shows the benefits of combining sociophonetic and social psychological approaches, and points out the value of applying novel methods and technologies. Beyond that, the present thesis offers useful implications for practice. Speech stereotypes in the context of sexual orientation can be rejected as inaccurate – for example, native German straight men do not nasalize more or less than gay men. According to EPSOM, the actual sexual orientation cannot be inferred from perceived sexual orientation. Thereby, the present thesis contributes to an erosion of stereotypes and a potential reduction of homonegative discrimination.

1 Introduction

Imagine you listen to an unfamiliar moderator in a radio program, hear an off-voice speaker commenting on a documentary or a news story, or receive a phone call from an unknown number. In all three cases, you have no clue about the person's appearance or about how she or he moves. You just hear a voice. And in all three cases the voices do not only invite you to process what is said (linguistic information), but also how it is said (paralinguistic information), and who said it (extralinguistic information). A voice allows you not only to discern a person's gender (mostly) correctly, but also to obtain information about her or his socio-geographic background (nationality, regional origin, ethnicity) and age. Gaydar research[1] has repeatedly shown that sexual orientation can also be rated with above chance accuracy on the basis of voice recordings, as well as of faces (see review by Tskhay & Rule, 2013). Recent studies suggest that perceivers' stereotypes about sexually divergent groups – which are mostly connected to masculinity/femininity – influence the identification of sexual orientation (Cox et al., 2015). Rating people as lesbian/gay or even bisexual irrespective of their actual sexual orientation has extensive and often negative consequences ranging from impolite service to physical violence (see meta-analysis by Katz-Wise & Hyde, 2012; Herek et al., 2002).

The superordinate aim of the present thesis is to promote the understanding of speech and appearance as social markers of sexual orientation and gender by combining producer- and perceiver-centered approaches. Particularly, the thesis sheds light on 1) the accuracy of speech stereotypes because of their behaviorally relevant social consequences and 2) how sexual orientation is interpersonally construed by developing a new model. The superordinate aim comprises some important differentiations (speech vs. appearance, sexual orientation vs. gender, producer-centered vs. perceiver-centered approaches) that need to be explicated in order to elucidate the prioritization of the present thesis.

[1] Gaydar is a compound of "gay" and "radar". Gaydar research deals with the question whether perceivers are able to identify targets' actual sexual orientation correctly.

Although sexual orientation and gender are stereotypically associated with each other, they are not as interchangeable as lay people's sexual orientation stereotypes might suggest (Kite & Deaux, 1987). This is especially true for gender-role self-concept (i.e. self-ascribed masculinity/femininity) in comparison to gender identity (see section "Gender-Role Self-Concept). Distinguishing both of these gender-related conceptualizations makes it possible to separate the biological and social effects of gender on speech (see Smyth & Rogers, 2008). In fact, this is exactly what the present thesis is primarily about: The main focus is on speech and its acoustic characteristics, because previous social psychological research was mostly concerned with faces as social markers of sexual orientation (for exceptions please see Rieger et al., 2010; Valentova & Havclíček, 2013; Sulpizio et al., 2015). Additionally, the present thesis builds on the rich evidence on single acoustic parameters attained by sociophonetic research and provides explanations for their inconsistent result patterns (for a review see Munson & Babel, 2007). Therefore, cross-fertilizing effects of social psychological and sociophonetic research are used to pursue the superordinate aim of the present thesis. Although this thesis sets a high value on acoustic characteristics, faces as special markers of appearances expand into the perception of sexual orientation. This points to the third differentiation: The present thesis centers not only on producers (e.g., how speakers express their sexual orientation via voices) but also on perceivers (e.g., what kind of vocal information do listeners use to derive a certain sexual orientation). In previous social psychological gaydar research, producer- and perceiver-centered approaches co-existed mostly disconnected from each other. Yet the integration of both approaches is important for designing a model on the interpersonal construction of sexual orientation and determining the accuracy of speech stereotypes.

In the first chapter, an overview about the theoretical and empirical foundations of the present research is provided. This overview begins with some remarks on the content and structure of stereotypes in the context of sexual orientation, on their measurement, and on different kinds of stereotype accuracy. Because sexual orientation stereotypes are generally linked to gender, the subsequent section provides information on gender-role self-concept, on how it differs from gender identity, on measurement problems occurring in 21st century, and on whether lesbian/gay and straight people do actually show gender-role self-concept differences, as suggested by lay gender inversion theories. Afterwards, the overview turns to speech and briefly introduces speech production phenomena and processes linked to gender. Finally, the overview refers to expression and perception of sexual orientation by emphasizing acoustic parameters. However, evidence from face research is also reported here. The information provided in the overview is necessary to elaborate the aims of the present thesis in more detail. The five manuscripts, which form the basis of this thesis, will be presented in summary in the second chapter. In the third chapter, an interrelated discussion of the results will be carried out and the thesis' strengths, limitations, and potentials for future research will be illuminated.

Before directly turning to the introductory overview, I am concerned with one point: The linguistic reference to sexual orientation and gender often holds pitfalls due to potential discrimination risks. In line with APA publication guidelines to reduce language bias (American Psychological Association, 2012), I chose not to use the terms „homosexual“ and „homosexuality“ and have used "lesbians" instead of "lesbian women" throughout. Because the terms "heterosexual" and "heterosexuality" would trigger their antonyms, I also prefer to speak of "straight" and "straightness". Although this thesis primarily deals with women and men, I attempt to linguistically represent gender diversity instead of implying gender binarity. For example, I have tried to avoid terms like "both gender" or "other gender". After these initial clarifications, let's begin with a look at stereotypes in the context of sexual orientation.

1.1 Sexual Orientation Stereotypes

"Gay men speak nasally!" – that would probably be the most frequent answer, even before mentioning the "high-pitched voice" typically imputed to gay men, if you were to conduct a little survey about speech stereotypes in the context of sexual orientation in an arbitrary street in Germany. In other countries, associations between gay men and nasality are less frequent (Mack, 2010; Chiang, 2003), if they exist at all (Piccolo, 2008; Panfili, 2011). Although, this reflects speech stereotypes' culture-specificity, hints towards cross-cultural speech stereotypes have also been found: Independent of respondents' nationalities, gay men were characterized as having a high-pitched voice (Panfili, 2011; Piccolo, 2008; Mack, 2010; Chiang, 2003), whereas straight men were thought of as having a low-pitched voice (Piccolo, 2008). Moreover, another commonly expressed stereotype regarding gay men is that they are believed to have a lisp (Piccolo, 2008; Mack, 2010; Panfili, 2011). In comparison, explicit speech stereotypes about sexually divergent women seem to be less prevalent (Jacobs, 1996; Munson et al., 2006). This possibly reflects the putative absence of a female equivalent of the gay voice (Zwicky, 1997). On the other hand, lesbians are stereotypically characterized as having a low-pitched voice and showing a flat tone (Queen, 1997; Piccolo, 2008), whereas straight women's voices are typically heard to be high-pitched (Piccolo, 2008).

Nasalizing gay men, high-pitched straight women, lesbians speaking with a flat tone – these are all stereotypes. Stereotypes can be defined as associations between certain features (e.g., nasalizing) and social groups (e.g., gay men; see Allport, 1954, for a similar definition). Stereotypes are organizing elements that structure our social knowledge about groups and their members. Because they are generalized representations of social group characteristics, they are assumed to be valid for every member of a given group (Judd & Park, 1993). Depending on their potential social relevance, stereotypes either reflect socially shared beliefs about social groups (i.e., consensual stereotypes that are expressed by an average across different people) or individual beliefs about social groups (i.e., personal stereotypes that expressed by a single value of one person; Jussim, 2012; Ryan, 2002). Because consensual beliefs are held by different individuals in a given culture or population, they inhere a higher potential to become socially meaningful than personal stereotypes that are not socially shared. Since they allow different individuals to cross-validate their knowledge about a social group, consensual stereotypes can also be assumed to be more rigid and longer lasting. Because of their greater importance, the present thesis will focus on consensual stereotypes (henceforth: stereotypes).

In general, stereotypes in the context of sexual orientation refer to the assumption of gender (non)conformity: Lesbians have been expected to move with shoulder swagger, to appear rough, and to have a cool gaze, while gay men are believed to walk with hip sway and to appear warm (Webbink, 1981). According to lay people's gender inversion theories (Kite & Deaux, 1987), gay men are assumed to be more similar to straight women than straight men are (effeminization/de-masculinization of gay men; Madon, 1997), whereas lesbians are believed to be more similar to straight men than straight women are (masculinization/de-feminization of lesbians). While lesbians and gay men are marked as deviating from gender-role norms and hence, to be construed as “the others”, straight women and men are implicitly accepted to be the guardians of a gender-appropriate standard. Thus, gender conforming characteristics are generally not in the focus of attention, not questioned, and not the effect to be explained (Bruckmüller, 2013). Most previous research on social markers of sexual orientation and gender can be understood against this backdrop. In order to provide a more balanced view on social makers of sexual orientation and gender, lay gender inversion theories will be re-framed in the present thesis and considered to be lay gender convergence theories (henceforth: “lay gender convergence theories” instead of “lay gender inversion theories”). By doing so, the present thesis centers, questions, and explains gender conformity and connected heteronormativity[2] and asks whether the corresponding speech stereotypes are actually accurate.

How might one test whether straight women and men sound “straighter” and more gender conforming than lesbians and gay men? Stereotype accuracy refers to the extent to which beliefs about group characteristics correspond to actual group characteristics (Jussim et al., 2015). Hence, studies providing relevant evidence on the assessment of stereotype accuracy require three steps (Jussim et al., 2015):

1. Assessment of people’s beliefs about group characteristics (e.g., are straight women believed to speak in a more feminine way?)
2. Identification of criteria reflecting group characteristics (e.g., analyzing acoustic parameters of speech recordings from lesbians and straight women)
3. Comparison of perceived and actual group characteristics

According to remarks by Judd and Park (1993), three different kinds of stereotype accuracy can be distinguished: Stereotypic, valence, and dispersion accuracy. Stereotypic accuracy is defined by a correspondence of perceived and actual group characteristics: Stereotypic group characteristics would not be overestimated and counterstereotypic group characteristics would not be underestimated but accurately reflect social reality instead of being caricatures or exaggerations of little kernels of truth (Prothro & Melikian, 1955; Allport, 1954). Hence, attention should be paid to the degree of correspondence between beliefs and reality. A possible finding that straight women exhibit more gender conforming

[2] Heteronormativity can be defined in short as an unexpressed expectation that all people are straight (Warner, 1993) and hence, is accompanied by prescribing straightness the norm everyone should adhere to.

acoustic parameters than lesbians can be regarded as accurate and inaccurate depending on the extent of the stereotypic belief. If straight women are believed to be somewhat more gender conforming than lesbians, a slight acoustic difference in the supposed direction could be taken for inferring stereotype accuracy (i.e., believed difference and real-world difference would match). However, if straight women are seen as guardians of a gender-appropriate standard whereas lesbians were believed to sound overall masculine, the same slight difference would not be enough to deduce stereotype accuracy but inaccuracy (i.e., the believed difference would be much stronger than the real-world difference).

Valence accuracy as a second kind of stereotype accuracy refers to the correspondence of perceived and actual evaluations of a group and their characteristics. In most cases, attributes have an evaluative tendency (e.g., gender conforming behavior is considered to be more positive by most people than gender non-conforming behavior). Thus, valence accuracy would not be given if groups and their characteristics were systematically judged as more negative or positive than they really are. Because the measurement of real-world negativity/positivity without referring to people's valences seems to be accompanied by important obstacles, valence accuracy will not be regarded in the present thesis.

A third type of stereotype accuracy is dispersion accuracy, which refers to a correspondence of perceived and actual group variability around its central tendency. As explicated above, generalizations can be considered as key elements of stereotypes. Consequently, dispersion accuracy would not be given if members of one group demonstrate varying characteristics because their dispersion would be underestimated. The present thesis has a special focus on stereotypic and dispersion accuracy as two components of stereotype accuracy.

Since the first step in the assessment of the accuracy of speech stereotypes is to assess people's beliefs about the speech characteristics of lesbians, gay men, straight women and men, the question arises how to determine them. Stereotypes can be determined using three approaches: applying explicit methods, utilizing implicit methods, or deriving stereotypes from lay gender convergence theories. Speech stereotypes in the context of sexual orientation have been explicitly determined by authors asking themselves introspectively (e.g., Lakoff, 1973; Queen, 1997) or by questioning others in qualitative interviews (e.g., Chiang, 2003; Piccolo, 2008; Mack, 2010; Panfili, 2011). Up to now, no quantitative study has been conducted. In contrast, implicit approaches used more indirect methods. For instance, speech characteristics such as single sounds or suprasegmental elements (e.g., voice pitch), that are assumed to be stereotypically relevant, were altered in recordings of natural speech by digitally manipulating them using the matched-guise technique (Mack & Munson, 2012). Thereby, the entities of interest (e.g., /s/-sounds in natural speech) were substituted by similar acoustic elements (e.g., correctly articulated /s/-tokens vs. misarticulated /s/-tokens produced by a trained phonetician) that differed systematically (see Mack & Munson, 2012). Because correctly articulated tokens of /s/ were rated as straighter than misarticulated /s/-tokens for male speakers, a stereotype of straight men correctly articulating /s/-sounds could be inferred. Another implicit method is to correlate raters' judgments about speakers' sexual orientations with speakers' acoustic parameters (e.g., Gaudio, 1994; Munson et al., 2006; see section "Perception of Sexual Orientation" for further information).

Neither explicit nor implicit, the third method is to infer assumptions about people's beliefs about speech characteristics of lesbian/gay and straight women and men from lay gender convergence theories. These theories suggest that straight women and men show gender conforming and heteronormative speech characteristics, and vice versa for lesbians and gay men. A precondition for inferring speech stereotypes regarding sexual orientation on the basis of lay gender convergence theories is knowledge about which speaking patterns can be understood as gender conforming. However, before elucidating acoustic gender differences (see section "Acoustic Differences Based on Gender"), the assumption should be corroborated that straight women and men are considered as the guardians of a gender-appropriate standard whereas lesbians and gay men are thought of as deviating from prescribed gender-role norms.

1.2 Gender-Role Self-Concept

Do lay gender convergence theories – proposing gender conforming characteristics for straight women and men and gender non-conforming characteristics for lesbians and gay men – possess a kernel of truth? If not, deductions of speech stereotypes from lay gender convergence theories would not be accurate and should be disregarded. Therefore, gender-role self-concept differences in sexually divergent people are illuminated in the present section.

Gender-role self-concept can be defined as the part of an individual's self-concept that contains information about internalized gender-related stereotypes. Hence, gender-role self-concept is about how individuals perceive themselves as representing traits, behaviors, interests, attitudes, and appearances that are typically referred to as masculine (e.g., self-reliant, fixing the car) and/or feminine (e.g., warm, caring about others). In contrast to gender identity that refers to a mental representation of one's biological gender, gender-role self-concept refers to self-ascribed masculinity and/or femininity.

Psychological gender research has established a variety of instruments to measure different facets of self-ascribed features that are stereotypically considered masculine/feminine, such as personality attributes, behaviors, interests, and attitudes (Beere, 1990). In the 1970s, the separation of masculinity and femininity as two independent constructs and the accompanied conceptualization of androgyny (Bem, 1974) caused a quantum leap in the development of gender-role self-concept measures (Beere, 1990). However, in recent years, existing attribute-related scales have not been able to display gender differences in self-ascribed masculinity/femininity in Western societies (e.g., Abele, 2003, Sczesny et al., 2004, Evers & Sieverding, 2014), which can arguably be explained by gender role changes during the last decades (e.g., Diekman & Eagly, 2000; Spence & Buckner, 2000; Wilde & Diekman, 2005; Ebert et al., 2014). Hence, existing scales seem no longer to reflect relevant aspects of women's and men's gender-role self-concepts.

Do straight women and men show different gender-role self-concepts from lesbians and gay men? Supporting lay gender convergence theories, in a meta-analysis (overall n = 9,273), straight women and men attributed more gender conforming interests to themselves and described themselves as more gender conforming on a global level of masculinity/femininity than lesbians and gay men did; gender-related attributes such as instrumentality and expressiveness revealed considerably smaller differences (Lippa, 2005). Moreover, in another meta-analysis of retrospective studies on gender-role self-concept during childhood, straight women and men remembered more gender conforming behavior than lesbians and gay men (Bailey & Zucker, 1995). This association could of course be due to selective recall and memory biases based on high self-stereotyping. But even when memory biases were ruled out using home-made videos of the participants during childhood, pre-straight individuals were rated as more gender conforming than pre-lesbian/gay participants by others (Rieger et al., 2008). Hence, the stereotypic accuracy of lay gender convergence theories receives certain support by these findings. However, a considerable intra-group variance should be taken into account (Bailey & Zucker, 1995) which can be illustrated best among lesbians who either classify themselves as “butch” (more masculine) or “femme” by trend (more feminine; Pearcey et al., 1996; also see Singh et al., 1999). Thus regarding dispersion, stereotypes seem to be less accurate. Consequently, sexual orientation and gender conformity are not as interchangeable as lay people’s gender convergence theories might suggest.

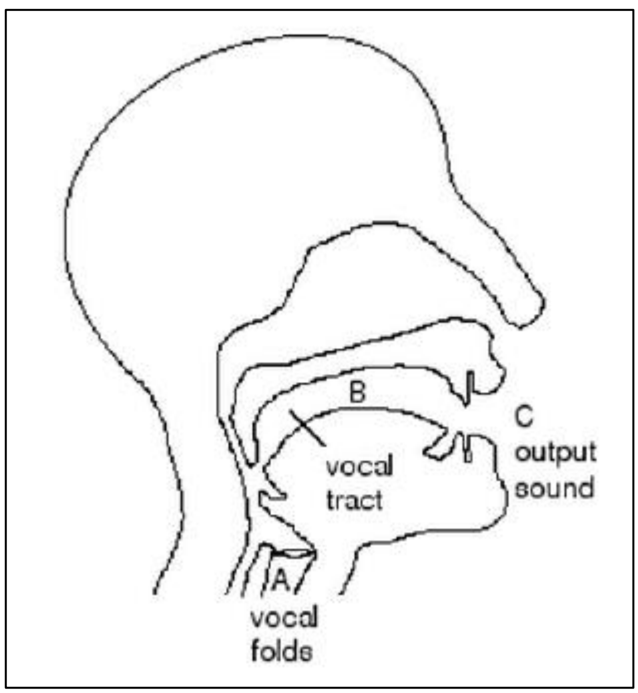
Taken together, nowadays, existing instruments for measuring self-ascribed masculinity/femininity are no longer able to validly and reliably represent gender-role self-concept in Western societies. This shortcoming is addressed in the present thesis. It is necessary to design a new gender-role self-concept scale in order to separate the effect of biological and social gender on speech. Despite considerable intra-group variance, existing gender-role self-concept measures provided support for stereotypic accuracy: Straight women and men described themselves as more gender conforming than lesbians and gay men did. Hence, it seems plausible to deduce stereotypic expectations about the speech of sexually divergent speakers from acoustic gender differences.

1.3 Acoustic Correlates of Gender

The following section provides an overview of how people’s speech differs according to their gender. This forms a foundation for understanding which acoustic parameters can be perceived as gender conforming and hence, are more likely to be shown by straight women and men. With reference to lay gender convergence theories, the perspective adopted in most previous sociophonetic studies on sexual orientation is illustrated. Moreover, this section gives information about speech fundamentals necessary for understanding the present research: What is meant by sociophonetics? How is speech produced and what are relevant acoustic parameters in the context of gender and – by deduction – of sexual orientation? And how to classify sociophonetic studies dealing with masculinity/femininity or sexual orientation?

Phonetics is the area of linguistics that is concerned with the production, transmission, and perception of spoken language. Sociophonetics as a discrete phonetic domain deals with how speech variation relates to social factors such as gender, age, or social class (Foulkes & Docherty, 2006) and is linked to the concept of socio-indexicality – how certain speech patterns are marked by socially relevant information (Foulkes & Docherty, 2006; Ochs, 1991). Hereby, a strong association to the social identity approach, often taken in social psychological research, becomes apparent. By combining social identity theory (Tajfel & Turner, 1979) and self-categorization theory (Turner et al., 1987), the social identity approach presumes that people classify themselves into certain social groups or categories along different axes of relevant social differentiation (gender, age, social class). This categorization serves an identity formative function by affirming the membership in one group and simultaneously negating the belonging to another group. Individuals' social identities are characterized by multiple belongings to groups on different social axes instead of a single group membership. However, members of the same social category are believed to have some characteristics in common that distinguish them from members of another social category. When a certain group identity is salient, people tend to behave as group members instead of idiosyncratic individuals (Hornsey, 2008). One branch of sociophonetic research tries to explain the behavioral consequences of social categorization in speech (e.g., how does a self-identification as lesbian/gay vs. straight affect speech production; Sulpizio et al., 2015), whereas another branch deals with social context effects on speech and identity performances (e.g., how do the same group of actors acoustically differ when playing a lesbian/gay or a straight role; Cartei & Reby, 2012). Because the social identity approach is applied in the present thesis, the discussion will focus on studies (implicitly) using the same approach.

The most frequently examined social factor in sociophonetics is speaker gender (e.g., Skuk, 2014). Speech processes such as articulation, phonation, and respiration are limited by biophysical inevitabilities (i.e., physiological configurations of the speech apparatus). Within these given physiological constraints, speakers' (un)conscious decisions about speech usage determine how speech is finally realized. These decisions are mainly co-determined by internalized norms about what is appropriate for a member of a particular social group in a certain situation and can sometimes result in exaggerations of purely physiological differences (see Jacobs, 1996; Jacobs et al., 2000). Hence, acoustic differences between women and men arise from both physiological and socially learned gender differences (e.g., Günzburger, 1989; Pierrehumbert et al., 2004; Simpson, 2009). However, there is disagreement regarding the extent to which physiological and social factors each influence acoustic gender differences (Munson, 2011). Moreover, most sociophonetic studies on gender implicitly examine biological and social effects of gender on speech in combination by comparing female and male speakers. Female and male speakers not only differ in their gender identity but also in their social gender aspects. Only a few studies exist which systematically examine acoustic correlates of masculinity/femininity as the social gender aspect. After presenting sociophonetic research on gender in general, studies on masculinity/femininity are explicated.



Note: Vocal folds (A) represent the source and vocal tract (B) represent filter of vocal sounds.

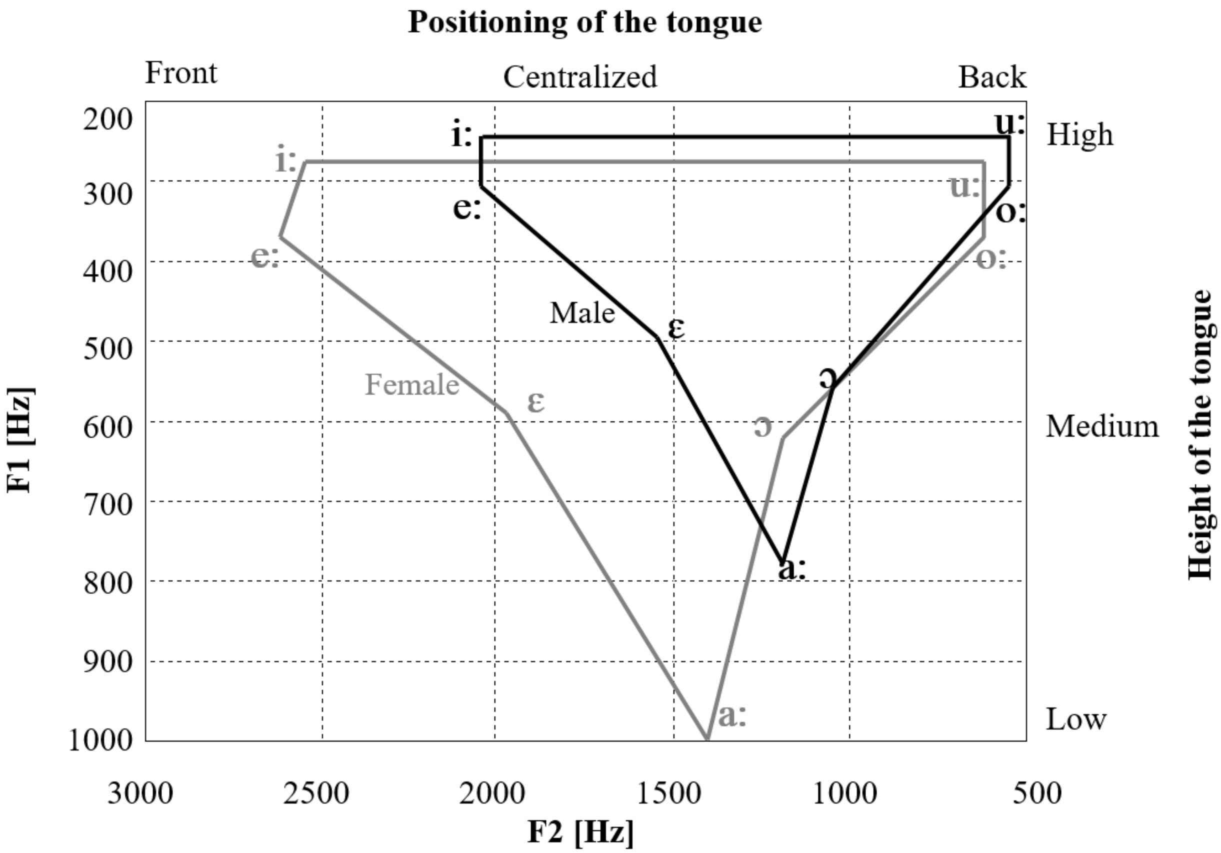
Figure 1. Source-Filter Model of Speech Production (Adapted from Kouroupetroglou & Chrysochoidis, 2014).

1.3.1 Gender Differences and a Crash Course in Speech Production

The following overview adapts and rearranges existing contributions on speech production and acoustic differences between female and male speakers in order to outline mostly acoustically relevant aspects for understanding the present thesis (e.g., Smyth & Rogers, 2008; Simpson, 2009; Skuk, 2014). The most meaningful biophysical gender differences result from sexual dimorphism during adolescence and are reflected in both components of the source-filter model of speech production (Fant, 1966; see Figure 1). This model proposes a mainly egressive air stream passing the larynx as the source of speech production. This air stream causes the vocal folds, which are located within the larynx, to vibrate as they are brought together. One vibration equals one cycle of vocal fold opening and closing. The frequency of vocal fold vibration per second is called fundamental frequency (henceforth: f0). In comparison to female speakers, the male larynx and vocal folds tend to become more massive during puberty, which reduces the readiness of the male vocal folds to vibrate. Male f0 averages 100-120 Hz, whereas female mean f0 is around 200-220 Hz in German (Simpson, 2009). Because mean f0 is directly linked to voice pitch as its perceptual correlate, male German voices sound lower-pitched than female German voices. However, the gender difference in mean f0 is more distinctive than can be expected based purely on laryngeal positioning and configuration (Günzburger, 1989; Pierrehumbert et al., 2004). For example, the mean f0 difference for women and men speaking the Chinese Wù dialect is only 17 Hz (see Simpson, 2009). Moreover, compared to female speakers, male speakers tend to utilize a smaller area of the available f0 range[3] (i.e., Jacobs et al., 2000), possibly as a means of vocal gender identity demarcation (McGonnell-Ginet, 1983). However, some studies found that differences between female and male speakers in f0 range are due to artefacts of speech signal analysis (e.g., Henton, 1989). Nevertheless, exaggerated gender differences in mean f0 and f0 range for German women and men seem to occur most likely in order to conform to gender-role norms (see Günzburger, 1989).

F0 can be considered as the most important acoustic parameter in vocally expressing gender (Skuk, 2014). However, when artificially eliminating f0 differences by electronic manipulation (e.g., applying an electrolarynx; Coleman, 1971) or when f0 is naturally absent

[3] F0 range is the difference between upper f0 boundary and lower f0 boundary and corresponds to voice pitch range.



Note: Acoustic vowel space formed by selected German vowels for female (grey) and male speakers (black).

Figure 2. Gender Differences in Acoustic Vowel Space (Adapted from Simpson & Ericsdotter, 2007).

as in whispered voices (Lass et al., 1976), listeners are still able to correctly identify female and male speakers. Besides f0, other acoustic characteristics related to the filter component of the source-filter model are relevant for determining speakers' gender (vgl. Smyth et al., 2003). The source signal produced in the larynx contains a spectrum of overtones that are integer f0 multiples and that are weakened or strengthened according to the configurations of the vocal tract (Skuk, 2014). The vocal tract corresponds to the supralaryngeal oral cavity ending at the speaker's lips and serves as a filter for the sound produced by the vocal folds. For example, the vocal tract settings cause distinctive vowel qualities (Simpson, 2009) that are primarily determined by the first two formant frequencies (Rendall et al., 2008). The first formant frequency (henceforth: F1) corresponds to the distance of the tongue dorsum to the hard palate, whereas the second formant frequency (henceforth: F2) corresponds to the horizontal positioning of the tongue in the mouth (see Figure 2). Using both dimensions, all vowel qualities of the German vowel system can be sufficiently described. Because of its greatest distance from the hard palate in relation to other vowels in the German vowel system, /a/ exhibits the highest F1. Because of its maximally frontal tongue positioning and spread lips during articulation, /i/ shows the highest F2. During puberty, the larynx lowers more strongly for male than female individuals (Pierrehumbert et al., 2004; Günzburger, 1989). Hence, the vocal tract for adult males (17-18cm) is longer than for adult females (14-14.5cm), which results in lower F1 in /a/ and F2 in /i/ for male than female speakers (Simpson, 2009). Consequently, the acoustic vowel space – the polygon formed by the distances between the single vowels in the two-dimensional acoustic space spanned by F1 and F2 – is on average less expanded and less shifted to the front and down for male than female speakers (Simpson & Ericsdotter, 2007; Munson & Babel, 2007; Pierrehumbert et al., 2004; see Figure 2). However, formant frequency differences between female and male speakers are higher than can be expected based on physiological vocal tract differences alone (Mattingly, 1966).

Vowel formant frequencies were found to be important signals to speakers' gender in the case of ambiguous mean f0 (Pernet & Belin, 2012). Besides f0 and formant frequency related characteristics, there are some additional acoustic parameters that reflect gender differences. In contrast to women, men have been found to realize the sibilants /s/ and /ʃ/ with lower peak frequency (Jongman et al., 2000), to use more creaky voice qualities (Henton & Bladon, 1988) and less breathy voice qualities (Henton & Bladon, 1985; Klatt & Klatt, 1990). Furthermore, gender differences in nasalance have been observed. Nasalance is defined as the proportion of the nasal sound to the overall nasal and oral sound and is linked to the perceptual impression of nasality. Male speakers exhibited lower nasalance than female speakers (van Lierde et al., 2001; Mishima et al., 2008).

Besides spectral acoustic parameters, temporal parameters also differ for female and male speakers. Contrary to the listener impression and stereotype of slowly speaking men, male speakers tend to exhibit an increased overall speaking rate (Byrd, 1992; Simpson, 1998) and produce shorter vowel durations (Simpson, 1998; Ericsdotter & Ericsson, 2001). Additional evidence was found when focusing on plosives. Plosives form a sound category that is defined by a total closure in the oral cavity produced by the tongue or the lips (given that the velum seals the nasal cavity) and a burst-like release of the oral closure a few milliseconds later. For the plosives /p t k/ male instead of female speakers exhibited a shorter duration between the burst-like release and the subsequent onset of vocal fold vibration (Whiteside & Irving 1997; Robb et al. 2005). Hence, several findings indicate the counter-stereotypical evidence that men tend to speak faster than women.

Taken together, there is a variety of differences regarding spectral and temporal parameters between female and male speakers caused by both biophysical inevitabilities and by socially learned gender-appropriate behavior. Based on lay gender convergence theories, straight men and lesbians can be assumed to reveal speaking patterns appropriate for a masculine gender role. This could be a lower mean f0 and f0 range, smaller acoustic vowel spaces, lower peak frequencies of sibilants, creakier and less breathy voices with lower nasalance, and higher tempo than straight women and gay men, who are believed to demonstrate speech appropriate for a feminine gender-role self-concept. However, stereotypes derived from lay gender convergence theories (e.g., low nasalance indexicalizing masculinity) do not necessarily correspond to stereotypes achieved by other methods. Gay men's voices in Germany are believed to be characterized by a special kind of nasality, namely hyponasality instead of hypernasality (for an illustration please see the German comedian Hape Kerkeling caricaturing a gay man[4]). Hyponasality comprises a decreased degree of nasalance and thus contradicts the prediction from lay gender convergence theories. Consequently, conclusions about speech stereotypes in the context of sexual orientation should be drawn with some caution from lay gender convergence theories and only in the absence of empirical sources for determining beliefs about others' speaking behavior.

1.3.2 Masculinity/Femininity and a Classification of Sociophonetic Studies

Although gender is the most often examined social factor in sociophonetics, a fine-grained analysis of gender in terms of actual masculinity/femininity that goes beyond a combined investigation of biological and social gender effects has not taken place (Smyth & Rogers, 2008). Actual masculinity/femininity does not only differ between but *within* female

[4] https://www.youtube.com/watch?v=RjrnlzXKBrA

and male speakers. Hence, when investigating female and male speakers separately and thereby holding biophysical gender differences of speech apparatuses constant, social effects of gender on speech can be analyzed in isolation.

Until 1990, being lesbian and gay was pathologized by the World Health Organization and considered criminal offense in many Western societies. Therefore, the willingness of lesbians and gay men to participate in sociophonetic studies on sexual orientation was limited, especially if they were asked to disclose their sexual orientation to an unfamiliar experimenter. Hence, until the investigation of masculinity/femininity became a research aim in its own right during the first decade of the 21st century (Biemans, 2000; Smyth et al., 2003), sociophonetic studies on masculinity/femininity constituted a mere vehicle for examining acoustic markers of sexual orientation. Sociophonetic studies on acoustic correlates of masculinity/femininity and sexual orientation can be classified as producer- and perceiver-centered studies. *Producer-centered studies* focus on the speaker as one actor within the communication process. They give information about the acoustic parameters speakers use to signal a certain identity or to take on a certain role. Hence, they are about actual masculinity/femininity (i.e., gender-role self-concept) or actual sexual orientation *Perceiver-centered studies* focus on the listener as the other actor within the communication process. They ask which acoustic parameters are associated with listeners' impressions of speakers sounding feminine or masculine and lesbian/gay or straight, respectively. Hence, listener-centered studies are a means to implicitly infer stereotypes regarding gender conforming and heteronormative speech because they deal with perceived masculinity/femininity and perceived sexual orientation.

To date, only three producer-centered studies have dealt with acoustic correlates of gender-role self-concept: One performativity-based producer-centered study (Andrews & Schmidt, 1997), one identity-based producer-centered study (Biemans, 2000), and one study with n = 2 (Chiang, 2003) which can be referred to as anecdotal evidence. Although the present thesis uses a social identity approach and thus, mainly focuses on identity-based studies, findings from Andrews and Schmidt (1997) are also included given the overall small number of producer-centered gender-role self-concept studies. *Performativity-based producer-centered studies* explore how speakers use different ways of speaking to perform different roles or to indicate the effects of different social contexts on speech. For that purpose, such studies contrast acoustic parameters taken from at least two recordings. Andrews and Schmidt (1997) compared the feminine and masculine self-presentations of the same group of eleven straight male cross-dressers. In line with lay gender convergence theories, in the masculine presentation mode, mean f0 was lower and overall speaking duration was shorter than in the feminine presentation mode. *Identity-based producer-centered studies* explore whether the speaking patterns of self-identified feminine and masculine or lesbian/gay and straight speakers, respectively, differ. Typically, the speech recordings of at least two different speaker groups are collected on one occasion. In her PhD thesis, Biemans (2000) investigated the voice quality correlates of gender identity and gender-role self-concept of 57 female and male speakers each. Gender-role self-concept was assessed using four instruments gathering data on feminine and masculine personality attributes and behaviors. Whereas almost all analyzed acoustic characteristics distinguished between female and male speakers, only four acoustic markers of gender-role self-concept were found for women and six for men. Because multiple significant tests were undertaken, alpha error was inflated. However, applying Bonferroni-correction formula would have reduced the number of significant correlations to zero. The small evidence can be attributed to the disputable validity of the gender-role self-concept measures. Three of four gender-role self-concept scales did not differentiate between female and male speakers and the scales showed a relatively high number of unexpected correlations (e.g., male speakers who self-reported more feminine personality attributes revealed higher loudness). This shortcoming additionally confirms the need for developing a valid gender-role self-concept measurement in order to provide relevant evidence for the acoustic correlates of actual masculinity/femininity.

In contrast, there are a number of perceiver-centered studies on masculinity/femininity (Terango, 1966; Gaudio, 1994; Avery & Liss, 1996; Biemans, 2000; Smyth & Rogers, 2003; Guzik, 2004; Munson, 2007). Only those acoustic parameters are summarized here that were investigated in at least two studies, because the explanatory power of studies on acoustic parameters that have been examined only once is limited. The only consistent difference was found for acoustic vowel space: More masculine sounding speakers produced a lower vowel frontalization in /i/ than less masculine sounding speakers (Avery & Liss, 1996; Munson, 2007). Some parameters were repeatedly shown not to be acoustic correlates of perceived masculinity/femininity: F0 range (Terango, 1966; Biemans, 2000), acoustic vowel space lowering (mean F1; Avery & Liss, 1996; Munson, 2007), and overall speaking rate (Terango, 1966; Avery & Liss, 1996; Biemans, 2000). For other acoustic parameters such as mean f0 (Terango, 1966; Munson, 2007; Biemans, 2000; Smyth & Rogers, 2008), f0 range (Terango, 1966; Biemans, 2000), center of gravity and skewness of /s/ (Avery & Liss, 1996; but see Munson, 2007) inconclusive results have been found.

Taken together, compared to sociophonetic studies on gender in general, only a few have dealt with masculinity/femininity. Regarding actual masculinity/femininity, existing identity-based studies suffered from methodological shortcomings (small sample size, questionable psychometric properties of gender-role self-concept measurement). Hence, no relevant evidence for gender-role self-concept effects on speech has been demonstrated. The present thesis aims to fill this gap. By doing so, a comparison of acoustic correlates of actual and perceived masculinity/femininity can be implemented that provides information on how speakers acoustically indexicalize their self-ascribed masculinity/femininity and on the acoustic parameters that perceivers typically associate with masculinity/femininity.

1.4 Expression of Sexual Orientation

A common finding in sociophonetic gender research is that speech differences within one gender group are considerably larger than differences between gender groups (Eckert & Podesva, 2011; see also Freed, 1995). Referring to intersectionality[5], individuals' social identities are constituted by multiple belongings to social groups because individuals possess different socially relevant characteristics that interact with each another. Generally speaking, genders are performatively construed in contrast to other individuals with the same gender, and not in contrast to other genders (Cameron, 2005). For example, working-class girls were found to pronounce /s/-sounds more similar to male speakers than to middle-class girls (Stuart-Smith, 2007). This is interpreted to suggest that working-class girls are more concerned with separating themselves acoustically from middle-class girls than from men. Hence, when dealing with social categories, the consideration of intersectionality is essential.

[5] Intersectionality is defined by Kirkshaw (2015) as „the idea that multiple axes of social differentiation intersect in producing systems of identification within a social matrix" (p. 629).

In the present thesis, gender and sexual orientation are examined intersectionally which seems to be especially advisable when taking a social identity approach. Because the previous section was dedicated to acoustics of gender, the present section provides an overview about how speaking patterns of lesbians and straight women as well as gay and straight men differ. By doing so, shortcomings of previous producer-centered studies are elaborated and the necessity for including new methodological approaches is emphasized (e.g., fine-grained psychological analysis).

Research on acoustic parameters used by speakers to signal their sexual orientation is a relatively recent enterprise (for early research, see Lerman & Damsté, 1969; Gaudio, 1994; Moonwomon-Baird, 1997). The same classification of producer-centered studies as has already been introduced for sociophonetic studies on acoustic correlates of gender-role self-concept, can be applied to acoustic studies on sexual orientation. However, performativity-based producer-centered studies will be disregarded in the present thesis because they either instructed speakers to record lesbian/gay or straight versions of the same text which results in stereotypical speech far removed from everyday speaking behavior (e.g., Crist, 1997; Cartei & Reby, 2012; Russell, 2015; Russel, 2017) or resemble case studies in that they recorded speech produced by one speaker in different situations which makes explanations referring to social identity aspects impossible (Podesva, 2006; Podesva, 2007; Podesva, 2011). Hence, keeping the social identity approach in mind, the focus of this thesis is on identity-based producer-centered studies that analyzed recorded speech provided by at least two speaker groups represented by more than one speaker each in their undisguised voice.

Previous studies on acoustic differences between lesbian/gay and straight women and men led to an overall inconsistent pattern of findings that will be exemplarily illustrated referring to mean f0. The finding that straight women realized a higher mean f0 than lesbians (Moonwomon-Baird, 1997; Camp, 2009; van Borsel et al., 2013) has not always been replicated (Munson et al., 2006; Rendall et al., 2008). The same is true for studies on men's speech: Whereas some have provided evidence that straight men speak lower-pitched than gay men (Linville, 1998; Baeck et al., 2011), the majority of studies have not (Lerman & Damsté, 1969; Gaudio, 1994; Sisson, 2003; Munson et al., 2006; Rendall et al. 2008; Zimman, 2010; Valentova & Havlíček, 2013; Sulpizio et al., 2015). For a more comprehensive overview containing all investigated acoustic parameters in the context of sexual orientation (e.g., f0 range, acoustic features of /s/, acoustic vowel space characteristics, voice quality parameters, and temporal measures) see Manuscripts 2 and 3 in the Appendix. To the best of my knowledge, no sociophonetic study on sexual orientation has investigated nasalance differences between lesbian/gay and straight women and men so far.

There are several reasons for the inconsistency of findings that are partially due to shortcomings of previous studies such as different languages investigated, small sample sizes resulting in low test power, different speech materials and methods for determining acoustic parameters, and different approaches to classify speakers' sexual orientation. However most importantly, a fine-grained analysis of speakers' psychological characteristics has been disregarded (but see Baeck et al., 2011; van Borsel et al., 2013). This seems remarkable because socialization-based explanations of speech differentiation in the context of sexual orientation unequivocally emphasize the importance of psychological characteristics.

First, when women adopt typical male speech and men adopt typical female speech this is assumed to cause lesbian/gay speech (Renn, 2003), whereby the adaptation of same-gender speech would lead to straight speech. Childhood and adolescence are supposed to form a first phase of acquiring sexually differentiated speech. Individuals who will identify as lesbian/gay are assumed to direct more attention to peers and adults who do not have the same gender (Renn, 2003; Pierrehumbert et al., 2004; Rendall et al., 2008; Smyth & Rogers, 2008), vice versa for individuals who identify as straight later in life. Hence, this explanation implicitly refers to lay gender convergence theories and gender-role self-concept in addition to social environmental characteristics.

Another explanation that does not replace the first one but complements it (Smyth & Rogers, 2008) is the assumption of a second acquisition phase in adulthood. According to this, lesbians and gay men pay attention to other lesbians and gay men who function as speech role models (Pierrehumbert et al., 2004; Smyth & Rogers, 2008) whereas straight women and men are geared to the speech of other straight people. Besides emphasizing social environmental characteristics, this explanation is implicitly connected to an increased psychosocial affiliation to the in-group (see Zwicky, 1997).

Whereas both the explanations above could be seen as general rules for speech acquisition affecting all individuals of a given group, two other assumptions directly refer to intra-group variability. One is the assumption that acoustic markers of lesbianism and gayness should be found only for lesbians and gay men who are comfortable with their sexual orientation and are open about it to others (Renn, 2003; van Borsel et al., 2013). However, this assumption implicitly sets acoustic markers of straightness as a standard and hence is not able to explain why straightness is learnt to be acoustically indexicalized (see statements on lay gender convergence theories in the section "Sexual Orientation Stereotypes").

Finally, a fourth assumption refers to the degree of sexual orientation causing differences in speech (van Borsel et al., 2013). Given a multiply graded sexual orientation measurement such as the 7-point Kinsey-scale (Kinsey et al., 1948), people exclusively self-identifying as lesbian/gay or straight are supposed to demonstrate speaking patterns different from people who show bisexual tendencies. However, the majority of previous studies have either given insufficient information on sexual orientation measurement or used a coarse grouping resulting in comparisons of straight vs. non-straight speakers (Lerman & Damsté, 1969; Gaudio, 1994; Sisson, 2003; Pierrehumbert et al., 2004, Munson et al., 2006, Rendall et al., 2008; Munson, 2010; but see Linville, 1998; Valentova & Havlíček, 2013; Sulpizio et al., 2015).

The present research intends to include a fine-grained psychological analysis accounting for psychological characteristics influencing the association of actual sexual orientation and speech (e.g., gender-role self-concept, social environmental features, psycho-social affiliation) and expected variability within sexual orientation groups. Given that previous studies have mostly disregarded a fine-grained psychological analysis, their inclusion can be regarded as applying novel methods to the research of acoustic correlates of sexual orientation. Connectedly, other techniques new to this field will be utilized such as voice morphing by the speech modification and resynthesis software TANDEM-STRAIGHT (Kawahara et al., 2008; Kawahara et al., 2009) and nasometry. Voice averaging as a special voice morphing technique serves the examination of stereotype accuracy and more directly the kernel of truth hypothesis (Prothro & Melikian, 1955). By averaging the acoustic parameters of the same utterance recorded from different speakers belonging to the same sexual orientation group, individual speech differences between speakers of the same group considered to be eliminated. Hence, voice morphing results in naturalistic sounding voice averages that mirror the acoustic characteristics typical of a given sexual orientation group. If voice averages from different sexual orientation groups sound alike, the kernel of truth hypothesis could be rejected. Nasometry is a procedure that helps researchers to determine speakers' nasalance and hence, is a prerequisite to investigate the stereotype accuracy of the allegedly nasalizing gay men. Importing new techniques is considered as beneficial for gaining new perspectives onto the field of acoustic correlates of sexual orientation.

1.5 Perception of Sexual Orientation

Perception in social psychological terms is accompanied by social categorization. People do not only categorize themselves, as has already been mentioned when introducing the social identity approach (see section “Acoustic Correlates of Gender”), but also others (Tajfel & Turner, 1979). Sorting other people into different categories is the first step in impression formation when encountering them for the first time, as has been explicated in the Continuum Model of Impression Formation (Fiske & Neuberg, 1990). In this initial phase, there is a strong influence of stereotypes on impression, because perception of a group membership is guided by the degree of similarity between the individual and the category prototype (Bodenhausen et al., 2012; Steffens et a., 2016). Thus, when the signals provided by an individual fit the stereotypes in the perceiver's mind and hence, are meaningful, categorization takes place (Trepte, 2006). Speech and appearance can provide meaningful information about the same category and have been shown several times to be used by perceivers in order to categorize others (see Rakić et al., 2011). Indeed, members of one group share certain characteristics which simplify navigation of the social world, because others do not necessarily have to be perceived and cognitively processed as individuals but as group members (Bodenhausen et al., 2012). To categorize others enables the perceivers to derive expectations in social domains which become behaviorally relevant for the perceivers in their interaction with the target (Trepte, 2006; Bodenhausen et al., 2012; Ambady et al., 2000).

Although biological gender, age, and socio-geographical aspects of one’s identity can be perceived almost perfectly well, other axes of social differentiation are perceptually more ambiguous, such as sexual orientation (Tskhay & Rule, 2013). Because sexual orientation detection is far from perfect – approximately 65% of participants are identified correctly (Tskhay & Rule, 2013) – straight people were taken for lesbian/gay and lesbian/gay people for straight. Hence, there is a strong suggestion to differentiate between actual and perceived sexual orientation, particularly in sociophonetic research that promotes the concept of lesbian/gay-sounding voices in contrast to lesbian/gay self-identification (e.g., Smyth & Rogers, 2008).

Yet despite these high rates of false ratings, sexual orientation has been repeatedly shown to be perceived with above chance accuracy based on both voice recordings and faces (e.g., Gaudio, 1994; Linville, 1998; Smyth et al., 2003; Munson et al., 2006; Rule & Ambady, 2008; Rule et al., 2008; Rule et al., 2009; Stern et al., 2013; Tabak & Zayas, 2013; Tskhay et al., 2013). However, there is little comparative research on which signal leads to a higher accuracy (but see Rieger et al., 2010; Valentova & Havlíček, 2013). In contrast to a lot of research on face-based gaydar conducted by social psychologists, only a small number of studies have dealt with the impact on sexual orientation perception of single facial characteristics measured using facial geometrics (Hughes & Bremme, 2011; Valentova et al., 2014). Perceiver-centered studies that ask raters to judge sexual orientation based on recordings of natural target voices and/or faces and that associate those judgments with signal inherent features can be used as a mean for determining implicit stereotypes. Different from face-based studies, sociophonetic research on sexual orientation stereotypes has provided a rich but inconclusive basis of evidence, especially for male speakers (for a detailed summary please see Manuscript 3).

As for producer-centered identity-based studies of sexual orientation, inconclusive results can be partially attributed to the omission of psychological target characteristics that are assumed to explain within-group differences (see also Waksler, 2001; Munson & Babel, 2007). For example, a gender non-conforming straight and a gender conforming gay target sample possibly account for why straight men have been rated as more feminine than gay men in one study (Valentova & Havlíček, 2013). Although psychological differences between lesbian/gay and straight women and men are assumed as constitutive for signaling sexual orientation (see section „Expression of Sexual Orientation“) their influence on facial or acoustic parameters has rarely been investigated (Biemans, 2010; Baeck et al., 2011; van Borsel et al., 2013). Hence, including a fine-grained psychological analysis of targets would help explaining inconsistent results.

Furthermore, there are several hints that impressions related to sexual orientation were decoded from signals as a precondition for inferring sexual orientation itself. As has already been mentioned (see section "Masculinity/Femininity and a Classification of Sociophonetic Studies"), acoustic parameters are used to judge targets' masculinity/femininity. In addition, when others were judged as more gender conforming, they were also judged as straighter by trend (e.g., Dunkle & Francis, 1990; Gaudio, 1994; Rieger et al., 2010; Freeman et al., 2010; Sulpizio et al., 2015). Besides, perceived gender-role conformity was shown to be a better predictor for actual sexual orientation than perceived sexual orientation (Valentova et al., 2014). Hence, when attempting to understand how sexual orientation is interpersonally construed, impressions related to sexual orientation should be accounted for in addition to a fine-grained psychological target analysis and producer- and perceiver-centered approaches need to be integrated.

1.6 Aims of the Present Research

The present research aims on a global level to expand the knowledge on social markers of sexual orientation and gender in speech and appearance by combining producer- and perceiver centered approaches. This superordinate aim can be divided into two major and two minor aims. While the major aims are indepdent goals in their own right, the minor aims additionally serve to achieve the major aims.

The first major aim is to scrutinize whether speech stereotypes in the context of sexual orientation are accurate for female and male speakers (*Major Research Aim 1*). Thereby, especially stereotypic and dispersion accuracy are focused. Do lesbian/gay and straight people differ in acoustic parameters? How do stereotypical beliefs about lesbian/gay and straight speech relate to their true speaking patterns? Are individuals belonging to one category that is actually homogenous, or is there evidence for within-group variability in speech? And is there something like a kernel of truth? Hence, the present thesis questions common speech stereotypes and bears the potential to explain and reduce them.

The second major aim is to provide and test a model that helps to understand how sexual orientation is interpersonally construed (*Major Research Aim 2*). It details that actual sexual orientation is accompanied by certain psychological differences (e.g., gender-role self-concept) that influence acoustic and facial signals. A perceiver who associates impression related to sexual orientation to those signals, finally arrives at a judgment of sexual orientation. Hereby, the expression (producer-centered) and perception (perceiver-centered) of sexual orientation will be interweaved in order to integrate previous findings from social psychological and sociophonetic research and to explain inconsistent findings.

The first minor aim is to separate the effects of biological and social gender on speech (see Smyth & Rogers, 2008) by examining acoustic correlates of gender-role self-concept in female and male speakers (*Minor Research Aim 1*). The effects of gender-role self-concept on speech constitute a link proposed by the model on expression and perception of sexual orientation. Hence, pursuing the first minor aim can be regarded as preliminary work for model development.

The second minor aim is to introduce novel techniques helping to promote sociophonetic research of sexual orientation such as a fine-grained psychological analysis, nasometry, and voice averaging (*Minor Research Aim 2*). All three methods are considered to improve the assessment of stereotype (in)accuracy. Moreover, the fine-grained psychological analysis is intended to clarify inconsistent result patterns of sociophonetic studies, to test assumptions on the acquisition of sexually divergent speech, and most importantly to be a precondition for designing the model on how sexual orientation is expressed and perceived. Thereby, a new scale will be created to validly and reliably measure gender-role self-concept. Altogether, these new methods, especially the fine-grained psychological analysis and voice averaging, will hopefully benefit the field of sociophonetics in general.

2 The Present Research

The present thesis comprises five manuscripts – all of them are empirical research papers – in order to reach the aims presented in the preceding section. In the present chapter, an overview will be presented for each manuscript. It will highlight how the manuscripts contribute towards pursuing the global aim and its corresponding major and minor aims (see Figure 3), the methods used for answering the research questions will be provided, the results of the studies will be outlined, and connections between the single manuscripts will be established in detail. For a list of manuscripts and the full-text manuscripts, please see the Appendix. The results of the different studies are presented in relation to each other in the chapter "General Discussion".

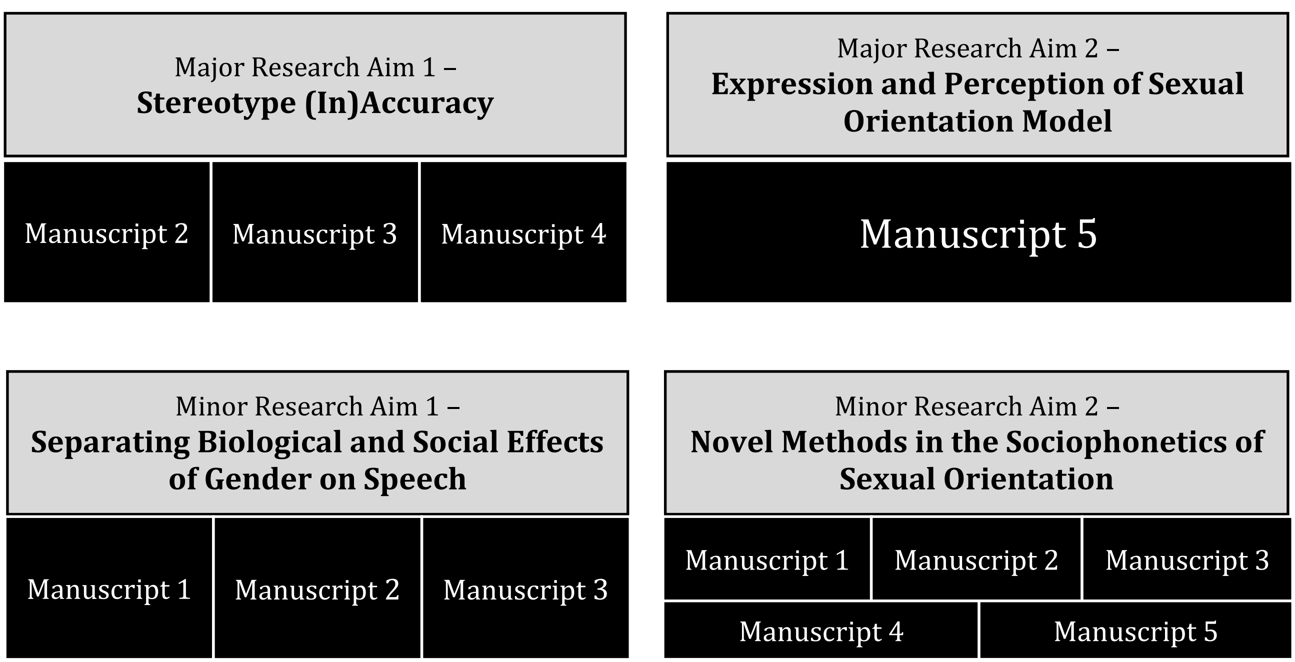


Figure 3. Relations of the Single Research Aims and Manuscripts.

Manuscript 1

Kachel, S., Steffens, M. C., & Niedlich, C. (2016).
Traditional Masculinity and Femininity: Validation of a New Scale Assessing Gender Roles.
Frontiers in Psychology, 7:956. DOI: 10.3389/fpsyg.2016.00956.

Against the background that existing instruments for measuring gender-role self-concept have not displayed gender differences in recent years (e.g., Abele, 2003, Sczesny et al., 2004, Evers & Sieverding, 2014), Manuscript 1 directly investigates an assumed higher-order masculinity/femininity construct by proposing the newly designed Traditional Masculinity-Femininity Scale (henceforth: TMF). The TMF was designed to represent overall, or "core", masculinity/femininity by reflecting different dimensions of gender-role self-concept such as physical appearance, behavior, interests, attitudes, and beliefs. The TMF comprises six items only (e.g., "Traditionally, my behavior would be considered as..."), which have to be rated on a 7-point scale ranging from 1 (*totally masculine*) to 7 (*totally feminine*). Providing a novel instrument for validly measuring actual gender-role self-concept is an essential precondition for introducing a fine-grained psychological analysis to sociophonetic research on sexual orientation (Minor Research Aim 2) and for separating biological and social effects of gender on speech (Minor Research Aim 1).

Overall, three studies have been conducted in order to determine the quality criteria of the whole scale and its single items. In all studies, reliability in terms of internal consistency, dimensionality, and convergent validity using a known-groups approach were tested. However, each study is characterized by a different individual focus. While in the Pilot Study female and male participants were contrasted regarding their TMF scores and the analysis primarily referred to item-specific quality criteria such as item homogeneities and item difficulties, Study 1 compared lesbians, bisexual and straight women and was mainly concerned with convergent validity. Study 2 was designed to replicate and expand Study 1's findings. Similar to Study 1, Study 2 dealt with convergent validity by using other gender-role self-concept-related measures (e.g., German Extended Personality Attributes Questionnaire, Runge et al., 1981). Going beyond Study 1, Study 2 contrasted women and men diverging in sexual orientation. Moreover, participants' f0 characteristics based on a short recording, the impression of straightness they provoked in others, and their contact to other lesbian/gay and straight women and men were analyzed in order to assess criterion validity. Finally, Study 2 dealt with a second measurement in order to provide data on TMF's test-retest-reliability and predictive validity.

Overall, the TMF was shown to parsimoniously and reliably measure (high internal consistencies and test-retest-reliability) a one-dimensional gender-role self-concept that differed between gender and sexual orientation groups in the expected directions (lower femininity/higher masculinity for men than women, for lesbians than bisexual and straight women, and for straight than gay men). Each item represented the scale very well and showed the expected difference between women and men (Pilot Study). TMF's convergent validity was indicated by moderate correlations of mean TMF scores with other gender-role self-concept-related scales and by providing clearer differences between lesbians vs. straight women and gay vs. straight men compared to other gender-role self-concept-related scales (Study 1 and 2). Moreover, evidence was provided for TMF's criterion validity: Higher gender conformity on TMF was accompanied by higher perceived straightness for women and men, more gender conforming f0 characteristics for women, and less contact to gay men for male participants. Additionally, TMF could be used to predict other gender-role self-concept-related measures one year later, which provides evidence for its predictive validity. Hence, the TMF is a suitable instrument for measuring gender-role self-concept and complements existing scales well.

The present manuscript forms the foundation for Manuscripts 2, 3 and 5, because these studies use the TMF as the main instrument for reliably and validly measuring actual gender-role self-concept in women (Manuscript 2), men (Manuscript 3), and both women and men (Manuscript 5). Because the present manuscript presents suggestive evidence that gender-role self-concept using TMF is acoustically reflected in speech (Minor Research Aim 1), it serves as preliminary work for a more encompassing investigation of the acoustic correlates of gender-role self-concept in women (Manuscript 2) and men (Manuscript 3). Connectedly, intra-group variability in lesbian/gay and straight women and men and its effects on speech can be reflected more adequately using the TMF, which is a precondition for examining the (in)accuracy of speech stereotype. Moreover, the present manuscript constitutes the groundwork for developing a model on the interpersonal construction of sexual orientation (Manuscript 5), because it suggests that differences in gender-role self-concept depending on actual sexual orientation are acoustically marked. Finally, the same pool of participants used in Study 2 of present manuscript served as a basis for all following manuscripts.

Manuscript 2

Kachel, S., Simpson, A. P., & Steffens, M. C. (2017).
Acoustic Correlates of Sexual Orientation and Gender-Role Self-Concept in Women's Speech.
Journal of the Acoustical Society of America, 141, 4793–4809. DOI*:* 10.1121/1.4988684

In contrast to sociophonetic studies on the acoustic correlates of male sexual orientation, there are only a few dealing with sexual orientation of female speakers (for a review, please see Munson & Babel, 2007). Evidence of previous studies is inconsistent in respect of f0 and acoustic vowel space characteristics, which is due to several reasons including an omission of psychological characteristics in most instances. In Manuscript 2, two studies are presented that are concerned with acoustic markers of actual sexual orientation and gender-role self-concept in women. Particularly, the present manuscript aims to scrutinize speech stereotypes about women's sexual orientation (Major Research Aim 1) and to separate the effects of biological and social gender on speech (Minor Research Aim 1) by implementing novel methodological approaches in sociophonetic research relating to a fine-grained psychological analysis (introducing gender-role self-concept and other psychological variables and applying a differentiated measure of actual sexual orientation; Minor Research Aim 2).

In both studies, women's sexual orientation was determined using a 7-point Kinsey-like scale ranging from 1 (exclusively lesbian) to 7 (exclusively straight; Minor Research Aim 2). Lesbians (Kinsey-like scores: 1-3) and straight women (Kinsey-like scores: 5-7) were recorded producing spontaneous and read speech. Their f0 and acoustic vowel space characteristics were examined using a speech signal analysis program. Most importantly, a fine-grained analysis of the speakers was done by collecting data on other psychological characteristics than sexual orientation such as gender-role self-concept, group affiliation, and social environment for all women, and coming-out measures additionally for lesbians. The

TMF as a newly designed tool for assessing gender-role self-concept was applied for separating the biological and social effects of gender on speech. Study 2 was intended to replicate the findings from Study 1 using another speaker sample and to expand the psychological data basis.

In both studies, lesbian and straight women differed not only in gender-role self-concept but also in all other psychological characteristics. Contradicting lay gender convergence theories and corresponding stereotypes, no acoustic differences between lesbians and straight women occurred. However, several aspects of intra-group variability were found: The exclusivity of sexual orientation on the Kinsey-like scale was acoustically indexicalized, there are acoustic markers of gender-role self-concept and other psychological characteristics at least for lesbians. Hence, the findings strongly suggest that lesbians and straight women could not be considered as acoustically homogeneous entities. Overall, the present manuscript suggests inaccuracy of the speech stereotypes towards women with regard to stereotypic content and group dispersion.

The pool of female participants in Study 2 in the present manuscript was the same as in Study 2 of Manuscript 1. However, in the present manuscript, bisexual people were excluded from analysis. Hence, the two samples used were similar but not equal. In contrast to Manuscript 1, in the present manuscript acoustic parameters were not used for scale validation but to investigate acoustic correlates of gender-role self-concept. In order to provide other speech data, the present manuscript was based on different acoustic analysis settings, selection of speech materials, and selection of acoustic parameters. Because the present manuscript showed the value of a fine-grained psychological analysis in sociophonetic research on sexual orientation in women, Manuscript 3 applied the same measures for a male speaker sample. Additionally, the recordings collected in preparation of the present manuscript served as a basis for the perception studies in Manuscripts 4 and 5. Finally, the present manuscript tested different paths of the Expression and Perceptions of Sexual Orientation Model (EPSOM) referred to in Manuscript 5: Associations of actual sexual orientation and psychological target characteristics, links between actual sexual orientation and implicit signals (i.e., voices), and connections of psychological target characteristics (e.g., gender-role self-concept) and implicit signals.

Manuscript 3

Kachel, S., Simpson, A. P., & Steffens, M. C. (2017).
Do I Sound Straight? – Acoustic Correlates of Actual and Perceived Sexual Orientation and Masculinity/Femininity in Men's Speech. In press in *Journal of Speech, Language, and Hearing Research.*

Findings on acoustic correlates of men's sexual orientation are similarly inconsistent as for women which can be attributed to the same reasons including the lack of a fine-grained psychological analysis. Manuscript 3 reports one study on the acoustic correlates of actual and perceived sexual orientation and gender-role self-concept in men. In the present manuscript, accuracy of speech stereotypes in the context of men's sexual orientation is scrutinized (Major Research Aim 1) by relating explicit and implicit speech stereotypes to acoustic differences depending on men's actual sexual orientation in a comprehensive literature review and by empirically testing them. Therefore, new methodological and technological approaches were introduced to sociophonetic research on men's sexual orientation such as fine-grained psychological analysis (which also serves to separate biological and social effects of gender on speech; Minor Research Aim 1) and nasometry (Minor Research Aim 2).

Gay (Kinsey-like scores: 1-2) and straight men (Kinsey-like scores: 6-7) were recorded producing spontaneous and read speech using a standard microphone on a first measurement and utilizing nasometry on a second measurement several months later. An extensive set of acoustic parameters was investigated using speech signal analysis software: F0 and acoustic vowel space characteristics, sibilant measures, plosive characteristics, and voice quality parameters such as nasalance. Speakers' sexual orientation was rated by a listener sample based on the same sentence for each speaker. Hence, implicit speech stereotypes were investigated by associating sexual orientation ratings and acoustic parameters. Moreover, speakers provided information about psychological characteristics (gender-role self-concept on the TMF, group affiliation, social environment, and coming-out measures for gay men only) and attributed different speech characteristics to gay and straight men for the assessment of explicit speech stereotypes.

Gay and straight men did not only differ in gender-role self-concept, but also in other psychological characteristics. The most important explicit speech stereotypes comprised voice pitch (low vs. high), nasality (non-nasal vs. nasal), chromaticity (dark vs. bright), and smoothness (hard vs. soft). According to implicit stereotypes, perceived straightness was associated with lower median f0, center of gravity in /s/, and mean F2, which is also in line with lay gender convergence theories. Explicit and implicit speech stereotypes only partially corresponded to each other, e.g., although there exists an explicitly labelled nasality stereotype, nasalance seems not to drive the perception of sexual orientation. Moreover, only one acoustic difference between gay and straight men was found (straight men showed lower mean F1 than gay men), which corresponded neither to explicit nor to implicit speech stereotypes. Incompatible with speech stereotypes, the exclusivity of sexual orientation on a Kinsey-like scale and actual masculinity/femininity were reflected in gay and straight men's speech. Hence, evidence for acoustically potent within-group variability was found. Consequently, differences within groups are more important than differences between them, and implicit and explicit speech stereotypes about gay and straight men seem not to contain a kernel of truth.

The pool of male participants was the same as in Study 2 of Manuscript 1. However, in the present manuscript, bisexual people were excluded from analysis. Hence, the two samples used are similar but not equal. The present manuscript is strongly linked to Manuscript 2 because both of them pursued the same aims. However, the present manuscript expanded the methodological approach of Manuscript 2 by empirically associating explicit and implicit speech stereotypes with actual speaking differences, by using nasometry, and by broadening the acoustic scope to other measures than f0 and acoustic vowel space characteristics. Moreover, the same stimulus sentence used for listeners' ratings of speakers' sexual orientation was utilized in Manuscript 4 for sexual orientation ratings by a different listener sample and hence, served as a basis for building the voice averages. Finally, the present manuscript tested different paths of the Expression and Perceptions of Sexual Orientation Model (EPSOM) presented in Manuscript 5 for men: Associations of actual sexual orientation and psychological target characteristics, links between actual sexual orientation and implicit signals (i.e., voices), and connections of psychological target characteristics (e.g., gender-role self-concept) and implicit signal.

Manuscript 4

Kachel, S., Radtke, A., Skuk, V. G., Zäske, R., Simpson, A. P., & Steffens, M. C. (2017). Do They All Speak the Same? – Investigating Sexual Orientation Information Using Voice Averages. Invited to Revise and Resubmit in *PLOS ONE*.

Manuscripts 2 and 3 suggest that speech stereotypes in the context of sexual orientation are inaccurate because they showed high acoustic within-group variability for lesbians, gay men, straight women and men. This suggests the existence of a variety of ways of acoustically reflecting actual sexual orientation. Furthermore, no group differences were found when relying on methodological approaches commonly used in sociophonetic research, namely aggregating acoustic parameters across a speaker sample. Taken together, Manuscripts 2 and 3 question if there is a kernel of truth in speech stereotypes. In the present manuscript, voice morphing as a novel technique (Minor Research Aim 2) is introduced to sociophonetic research on sexual orientation in order to more directly test the kernel of truth hypothesis (Major Research Aim 1).

Using voice averaging as a special voice morphing technique, the same utterances recorded from different speakers showing the same sexual orientation were averaged. Twelve voice averages were created that differed in gender (women vs. men), sexual orientation (lesbian/gay vs. bisexual vs. straight), and rating basis of sexual orientation (sexual orientation rated by the speakers themselves vs. rated by others). Each voice average was based on five speakers. For example, the same sentence that was uttered by five women who were rated as lesbian by a group of listeners was averaged. By applying voice averaging, individual differences between speakers of the same sexual orientation group are considered to be eliminated. Thus, voice morphing resulted in naturalistic sounding voice averages that mirrored the acoustic characteristics typical of a given sexual orientation group. Subsequently, a group of listeners judged the voice averages on sexual orientation in order to provide data on whether the voice averages still contain discernible sexual orientation information.

Voice averages of straight speakers were perceived as straighter than those of bisexual speakers who, in turn, were perceived as straighter compared to voice averages of lesbian/gay speakers, irrespective of speaker gender and whether speaker sexual orientation was self-rated or had been rated by others. Hence, sexual orientation information can be derived from voice averages, which indicates that speakers belonging to one sexual orientation group share a set of acoustic characteristics that listeners use to judge them. However, our findings also suggest that group stereotypes were exaggerations of existing between-group differences: Lesbian/gay voice averages based on other-ratings were perceived as much more lesbian/gay than those based on self-ratings, whereas there was no difference due to rating basis for bisexual and straight voice averages. Taken together, there is evidence that speech stereotypes are exaggerated versions of true differences, corresponding to a kernel of truth.

The present manuscript resorted to the same pool of speakers as Study 2 of Manuscript 1, Study 2 in Manuscript 2, and Manuscript 3. In contrast to the previous manuscripts, only five speakers per group were selected from the speaker pool for voice averaging and bisexual speakers were also considered. Moreover, the sexual orientation ratings by others, that were used to create half of the voice averages, are those reported in Manuscript 5.

Manuscript 5

Kachel, S., Steffens, M. C., & Simpson, A. P. (2017).
The Expression and Perception of Sexual Orientation Model: Speech Based Evidence.
Submitted to *Journal of Personality and Social Psychology.*

Manuscript 5 pursues the aim to conceptualize a model that explains how sexual orientation is interpersonally constructed (Major Research Aim 2). The Expression and Perception of Sexual Orientation Model (EPSOM) is based on different research lines in sociophonetics of sexual orientation and social psychology of gaydar and attempts to integrate them in order to explain inconsistent result patterns. EPSOM comprises five components that form an indirect route for transmitting actual to perceived sexual orientation. The indirect route proposes that *actual sexual orientation* influences other *psychological characteristics* of the targets (e.g., gender-role self-concept) which are reflected in certain configurations of *implicit signals* (e.g., mean f0 of voices or symmetry of faces). Raters decode these implicit signals in order to derive *impressions related to sexual orientation* (e.g., perceived gender-role self-concept) that are in turn used to finally arrive at *perceived sexual orientation*. EPSOM is deduced from a literature review that integratively presents existing connections between the single components. In two experiments, open research questions regarding the single components were addressed (e.g., is voice or face a better signal to sexual orientation?) and predictions of EPSOM regarding the indirect route were empirically tested.

In Experiment 1 a group of raters were asked to assess targets' sexual orientation based on speech recordings, facial photographs, and a combination of both. In Experiment 2, another group of raters judged targets' masculinity/femininity. Because the experiments were based on the same target pool already reported in Manuscripts 1, 2, and 3, both data on psychological target characteristics (e.g., gender-role self-concept) and acoustic parameters were available. A detailed analysis of features inherent in implicit signals was conducted for targets' speech recordings but not for their facial photographs in order to test for psychological target characteristics, implicit signals, and impressions related to sexual orientation mediating the effect of actual on perceived sexual orientation.

Several pieces of evidence were found for the indirect route proposed by EPSOM. Most importantly, the effect of actual on perceived sexual orientation is transmitted by the three proposed mediating components: Self-reported straightness led to a more gender conforming gender-role self-concept that provoked more gender conforming speech patterns which were perceived as rather gender conforming and finally effected an impression as straight. However, this was restricted to female targets only. Given the richness of the data, open research questions derived from previous literature are answered. For example, voices and faces were both shown to be suitable for judging targets' sexual orientation with above chance accuracy, but voices were a worse signal than faces. Taken together, Manuscript 5 showed how sexual orientation is interpersonally constructed by integrating producer and perceiver perspectives in EPSOM.

The pool of targets was the same as for all previous manuscripts. In the present manuscript, the same speech stimuli were used for male targets as in Manuscript 3. However, in contrast to Manuscript 3, the acoustic analysis in the present manuscript directly referred to the speech stimuli in order to establish a direct association between acoustic parameters of the utterances heard by the raters and the sexual orientation judgments that are based on these utterances.

3 General Discussion

The global aim of the present thesis was to expand the knowledge on social markers of sexual orientation and gender in speech and appearance by combining producer- and perceiver-centered approaches. This global aim was decomposed into two major and two minor research aims. The achievement of those aims and aim-related limitations are discussed in the “Locating the Results” section. Afterwards, the overall strengths and limitations of the present research are focused and suggestions for future research are presented.

3.1 Locating the Results

3.1.1 Stereotype (In)Accuracy

The first major aim of the present thesis was to scrutinize (in)accuracy of speech stereotypes regarding female and male sexual orientation in Germany. Speech stereotypes can be determined by different means and partially vary according to the method used, as can be seen regarding putative nasality differences between gay and straight men (Manuscript 3). Gay men were believed to produce a hyponasalized speech, while straight men were believed not to do so, when explicitly asking German raters. When applying implicit measurement, nasalance scores of speakers and listener ratings of sexual orientation of those speakers were not associated. This would suggest that speech stereotypes on male sexual orientation does not include nasality. In contrast, deductions from female-male speech differences according to lay gender convergence theories characterizing gay men as displaying hypernasalized speech compared to straight men. Thus, the content of some stereotypes depends on the selected methods to a similar degree as on the investigations’ socio-geographical context (see section “Sexual Orientation Stereotypes”).

Although speech stereotypes in Germany can be demonstrated using some methods, their real-life foundation seems to be small in any case. For male speakers, only one acoustic difference was found depending on sexual orientation – straight men were shown to produce a lower mean F1 compared to gay men (Manuscript 3) – whereas no speech differences between lesbians and straight women were found in any of the two female samples (Manuscript 2). No empirical evidence was demonstrated for two other prominent stereotypes: Neither for the explicitly expressed nasality stereotype for sexually divergent men, which that is especially prominent in Germany, nor for the cross-culturally shared voice pitch stereotype encompassing lesbians, gay men, straight women and men (Chiang, 2003; Piccolo, 2008; Mack, 2010; Panfili, 2011). This is in line with the general tendency of previous studies (see Manuscript 3 for a review on men) identifying relatively few speech differences between lesbian/gay and straight women and men. Hence, comparisons of acoustic parameters of lesbian/gay and straight women and men suggest stereotypic inaccuracy as one aspect of stereotype inaccuracy.

When focusing on dispersion inaccuracy as a second aspect of stereotype inaccuracy, the assumption was tested that members belonging to the same sexual orientation group are acoustically homogenous. Incompatible with this assumption, the present research has shown several pieces of evidence for within-group variability in speech for lesbians and straight women (Manuscript 2) as well as gay and straight men (Manuscript 3). In all sexual orientation groups, exclusivity of sexual orientation was acoustically marked, and all sexual orientation groups except for straight women acoustically reflected gender-role self-concept. Null findings of previous research on the acoustic indexicalization of variability within sexual orientation groups (Baeck et al., 2011; van Borsel, 2011) can be partially attributed to a restricted selection of acoustic parameters (only f0 characteristics were observed) and to a limitation of testing the variability for lesbians and gay men only. Hence, the present research suggests that sexual orientation groups cannot be treated as acoustically homogenous entities, which contradicts the generalizing character of stereotypes and further supports speech stereotype inaccuracy.

When using voice averaging as a novel technique in sociophonetic research on sexual orientation in order to additionally check for dispersion accuracy, speech stereotypes seem to comprise a kernel of truth (Manuscript 4). Averaging voices of different people belonging to one sexual orientation group is considered to eliminate individual acoustic differences while retaining acoustic characteristics typical for most or all members of the same sexual orientation group. Indeed, voice averages contained discernible sexual orientation information, because listeners judged straight voice averages as straighter than bisexual voice averages which were, in turn, rated as straighter than lesbian/gay voice averages, independent of speakers' gender. This finding supports the kernel of truth hypothesis (Prothro & Melikian, 1955), but it should be treated with caution. In order to maximize the chance for finding differences in voice averages, very distinctive sexual orientation groups were used (e.g., voice averages of speakers who judged themselves as exclusively lesbian/gay were contrasted with voice averages of speakers who judged themselves as exclusively straight). Hence, intra-group variability was limited a priori by speaker selection, which increased the chance for finding support for dispersion accuracy. Additionally, each voice average was based on five speakers only. Accordingly, it is not clear whether salient individual acoustic parameters have been entirely ruled out and hence, were not reflected in the voice average. Furthermore, the finding that voice averages based on sexual orientation ratings of others contained more sexual orientation information than voice averages based on speaker's self-identified sexual orientation can explain speech stereotypes as exaggerations of small speech differences and further supports stereotypic inaccuracy. Thus, although negligible evidence for accuracy of speech stereotypes has been found overall, small and exaggerated kernels of truth may serve as an explanation for the existence of speech stereotypes (Allport, 1954). Future studies using voice averages are recommended to use less extreme and more internally variable sexual orientation groups and to base voice averages on more than five speakers.

3.1.2 Expression and Perception of Sexual Orientation Model

The second major aim was to provide a model on the expression and perception of sexual orientation (EPSOM). EPSOM explains how sexual orientation is interpersonally construed and accounts for inconsistent results of previous research by integrating producer- and perceiver-centered approaches, by considering sociophonetic and social psychological knowledge, and by including a fine-grained psychological analysis expressed in the *psychological characteristics* component.

In the absence of explicit signals (e.g., rainbow flags), EPSOM's explanatory approach by proposing an indirect route can be associated with the inaccuracy of speech stereotypes (Major Research Aim 1). For example, straight women are believed to sound straight (see Rogers et al., 2001). In contrast, EPSOM provides a more nuanced perspective: To be perceived as straight is not only a matter of actual sexual orientation, but of gender conforming self-description, gender conforming speech, and gender conforming impressions provoked in the mind of the perceiver. Sexual orientation information has to pass the three mediating components proposed by EPSOM (*psychological characteristics*, *implicit signals*, and *impressions related to sexual orientation),* which can be considered as switching points on the way from actual to perceived sexual orientation. A straighter self-identification is accompanied by a more gender-conform self-concept, which leads to more gender-conforming self-presentation in terms of acoustic parameters. Based on the more gender-conforming speech, listeners derive more gender-conforming impressions that finally cause a straighter perception. By proposing an indirect route, EPSOM summarizes previous research from sociophonetics and social psychology. Although the present research has shown that three mediating components transmit the sexual orientation information, these components can simultaneously explain a loss of sexual orientation information. For example, it has been shown that there is considerable variance in self-ascribed masculinity/femininity for sexual orientation groups (Manuscript 1). Hence, there are some straight men who described themselves as rather gender ambiguous. According to the EPSOM logic, these less gender conforming straight men would provide less gender conforming speech. They would then be perceived as less gender conforming and thereby also as less straight/more gay. This would explain contra-intuitive findings (see Valentova & Havlíček, 2013). Consequently, relying on the mediating components, EPSOM explains why sexual orientation is perceived with above-chance but far-away-from-perfect accuracy, or in other words: why not all straight women were perceived as straight, and vice versa for lesbians.

Although EPSOM accounts for the inaccuracy of speech stereotypes by proposing an indirect route from producer to perceiver, it does not consider that speech stereotypes in the listeners' minds – independent of their (in)accuracy – influence the signal processing. Hence, EPSOM's indirect route corresponds to a rather bottom-up than top-down-processing and disregards that stereotypes may also influence low-level perceptual processes (Strand, 1999). Future research should account for top-down-processing and test for possible extensions of EPSOM.

Although there has been some evidence that voices were more revealing for male compared to female targets, confirmatory support for EPSOM's indirect route has been found for female targets only (Manuscript 5). This could be due to a restricted set of analyzed acoustic parameters in Manuscript 5 (i.e., f0 and acoustic vowel space characteristics). Although faces and other implicit signals are assumed to provide sexual orientation information, only speech-based evidence was presented. Moreover, the causality of EPSOM's indirect route has not been demonstrated because a regression-based approach was used for testing mediation hypotheses (Hayes, 2013). Consequently, future research needs to include more acoustic parameters, check for face-inherent features, and test for causality.

3.1.3 Separating Biological and Social Effects of Gender on Speech

The first minor research aim was to separate gender identity (e.g., female vs. male) and gender-role self-concept effects (e.g., feminine vs. masculine) on speech (see Smyth & Rogers, 2008). To make sure to overcome the shortcomings of the few previous studies on acoustic correlates of actual masculinity/femininity (Chiang, 2003; Biemans, 2000), the TMF was created as a tool for a fine-grained psychological analysis and was confirmed to be a valid and reliable instrument for measuring gender-role self-concept in women and men (Manuscript 1). Moreover, comparatively large-scale samples of female (n = 108; Manuscript 2) and male speakers (n = 54; Manuscript 3) were investigated separately: While gender identity effects were held constant, effects of actual masculinity/femininity on acoustics were observed. Hence, more and less gender conforming women and men were examined.

In line with lay gender convergence theories, acoustic correlates were found for all sexual orientation groups except for straight women: Lesbians who described themselves as more feminine, produced higher median f0; gay and straight men who called themselves more masculine, presented lower mean F2. To the best of my knowledge, this is the first investigation providing relevant data on the acoustics of gender-role self-concept.

When comparing acoustic correlates of gender-role self-concept (Manuscripts 2 and 3) and of perceived masculinity/femininity (Manuscript 5), the acoustic parameters indexicalizing actual masculinity/femininity were also used for the perception of masculinity/femininity. For female speakers, perceivers relied on mean f0 as a corresponding central tendency measure of the f0 distribution when determining masculinity/femininity indeed, but mean F2 was similarly important. For male speakers, perceivers based their masculinity/femininity judgments actually on F2 in /aː/ as another measure of vowel frontalization, but the speakers' lower fundamental frequency boundary was shown to be twice as important. Thus, findings suggest that perceivers assort additional acoustic parameters that are not primarily important in signaling gender-role self-concept when rating masculinity/femininity. Limiting this inference, acoustic parameters of actual and perceived masculinity/femininity are based on different speech materials and signal analysis procedures.

Nevertheless, investigating acoustic correlates of actual and perceived masculinity/femininity was a necessary precondition for establishing the EPSOM, because it arranged the paths from *psychological characteristics* to *implicit signals* and from *implicit signals* to *impressions related to sexual orientation*. Otherwise, EPSOM's mediations of the effect of actual on perceived sexual orientation by *psychological characteristics* (e.g., gender-role self-concept), *implicit signals*, and *impressions related to sexual orientation* (e.g., perceived masculinity/femininity) were empirically less well-grounded.

3.1.4 Novel Methods in the Sociophonetics of Sexual Orientation

The second minor research aim of the present thesis was to apply new methodological and technological procedures to sociophonetic research on sexual orientation in order to answer relevant questions. All of the three methods applied (nasometry, voice morphing, and a fine-grained psychological analysis), were used to assess the accuracy of speech stereotypes. By utilizing nasometry, the explicitly expressed stereotype of the "nasalizing gay men" was tested for a real-world foundation. Going beyond aggregated acoustic parameters for different groups, voice averaging as a special kind of voice morphing was used to examine the kernel of truth hypothesis more directly and was applied on whole sentences instead of single syllables or words for the first time. Most relevantly, implementing a fine-grained psychological analysis enabled the investigation of variability within sexual orientation groups questioning the dispersion accuracy of speech stereotype (Manuscripts 2 and 3).

Using a fine-grained psychological analysis allowed to further insights in the acoustics of sexual orientation and masculinity/femininity. The Traditional Masculinity-Femininity Scale as a tool for valid and reliable measurement of gender-role self-concept (Manuscript 1) permits an investigation of its acoustics in women (Manuscript 2) and men (Manuscript 3). Additionally, the inclusion of a fine-grained psychological analysis carries the potential to explain why previous studies found inconsistent acoustic correlates of sexual orientation. According to EPSOM, those studies that had shown straight men to produce a lower mean f0 than gay men (Linville, 1998; Baeck et al., 2011) had possibly taken gender conforming straight men and gender non-conforming gay men with a low intra-group variation each.

By the means of a fine-grained psychological analysis, also socialization-based assumptions for acquiring sexually divergent speech were tested empirically. The present research supports the assumption that exclusivity of sexual orientation is acoustically reflected (Baeck et al., 2011; van Borsel et al., 2013): The more exclusively straight straight women and men described themselves and the less exclusively lesbian/gay lesbians and gay men described themselves, the more likely they were to present gender conforming and heteronormative speech characteristics. While no speech differences between sexual orientation groups were found, the findings suggest that sexual orientation has an influence on speech within sexual orientation groups and is acquired in comparison to other members of the same group than other groups (see Eckert & Podesva, 2011 for gender-related speech).

Further socialization-based assumptions were tested for lesbians and gay men only and findings indicate that gay men and lesbians acquire sexual orientation related speech differently. Gay men who reported lower contact to girls during childhood, produced more gender conforming and heteronormative speech. This indicates an early acquisition of straight speech by gay men who disregarded girls as role models (see Smyth & Rogers, 2008). However, self-stereotyping memory biases could have co-determined these findings. Estimating the influence of adult role models on sexually divergent speech as another possible impact factor in childhood is recommended to be tested in future studies.

Furthermore, the more discrimination experiences gay men reported, the more gender conforming and heteronormative speech they produced. Although no causality can be inferred from correlations, the data can be interpreted in terms that gay men produce straight speech in order to prevent from discrimination experiences, because gay speech is socially sanctioned. In line with previous research (Baeck et al., 2011; van Borsel et al., 2013), we found no convincing evidence that lesbian/gay people who were more comfortable with their sexual orientation – indicated by coming-out age and extent for example – acoustically differed from those who did not.

In contrast to gay men, assumptions for speech acquisition in lesbians center on role models during adulthood. Lesbians showed a more gender conforming and heteronormative speech, the less current contact to other lesbians they reported, the higher they were affiliated to straight women, and the more straight friends they had. In contrast to highlight childhood for gay men, results strongly indicate the importance of adulthood for lesbians. Future studies are recommended to test assumptions for the acquirement of sexual orientation related speech in straight women and men.

3.2 Strength of the Present Research

The present thesis stands at the crossroads of sociophonetic and social psychological research and combines both approaches by applying a detailed investigation of characteristics inherent in voice signals and a fine-grained psychological analysis. Most previous (social) psychological studies, which mainly dealt with faces, did not attribute the expression and perception of sexual orientation to specific facial features but accessed faces on a rather global level (except for Rule et al., 2008; Rule et al., 2009; Hughes & Bremme, 2011; Valentova et al., 2014; Skorska et al., 2015). In contrast, sociophonetic research has a long-lasting tradition to consider signal inherent characteristics and provided a rich basis of possible acoustic parameters being used for expression and/or perception of sexual orientation. Although different assumptions have been formulated on the development of sexually divergent speech, sociophonetic research lacked in a fine-grained psychological analysis. By utilizing psychological instruments for measuring gender-role self-concept, social environment characteristics, group affiliation, and coming-out measures, and by a nuanced assessment of sexual orientation, the present thesis accounted for variability within sexual orientation groups. Thereby, an evaluation of (in)accuracy of speech stereotypes, a separation of biological and social gender effects of speech, and a test of assumptions on the acquisition of sexually divergent speech were enabled.

In contrast to most signal-related social psychological research on sexual orientation, the present thesis combines producer- and perceiver-centered approaches. This was a precondition for developing a model on the interpersonal construction of sexual orientation and for assessing (in)accuracy of speech stereotypes by comparing vocal signals used to express and perceive sexual orientation. For example, the present thesis expands research on speech stereotypes by a quantitative approach complementing qualitative studies. Using a quantitative approach enabled the present research to provide information about the strength of stereotypes instead of naming the number of perceivers carrying the speech stereotype. Because the quantitative approach was used for male participants in the present research only, future studies should implement them on sexual orientation in women.

Compared to previous research, the present thesis attempted to avoid making lesbians and gay men the effect to be explained (see Bruckmüller, 2013) when dealing with signals related to sexual orientation. Hence, lay gender inversion theories (Kite & Deaux, 1987) were re-framed as lay gender convergence theories. Thereby, the need of explanation was inverted: When dealing with sexual orientation it is probably even more necessary to elucidate how straightness and heteronormativity are established, performed, and maintained. Explicating how some lesbians and gay men transgress gender-conformity-related norms can give us hints for understanding deviations from heteronormativity but does not focus on heteronormativity. Hence, it would be better to switch perspectives and illustrate how straight people differ from lesbian/gay people in order to illuminate the unwritten rules in society prescribing how to behave.

3.3 Limitations of the Present Research and Future Directions

The present thesis is based on categorizations of self and others regarding sexual orientation and gender and hence, has taken a social identity approach. Thereby, this thesis differs from a branch of sociophonetic research that is concerned with speech performance of sexual orientation. Asking speakers to deliberately adopt speech considered as lesbian/gay or straight (e.g., Crist, 1997; Cartei & Reby, 2012) could be understood as another method of identifying speech stereotypes because highly artificial speaking behavior geared to beliefs about group-based speech were investigated. Moreover, recording a single speaker over different everyday situations (e.g., Podesva, 2006; Podesva, 2007) would have shed light on the contextual influences on speech but sexual orientation could not have been investigated because a single speaker is characterized by more than a certain sexual orientation. However, the examination of contextual effects on speech seems to be a valuable undertaking, because 1) natural and everyday surroundings would elicit speech with a higher ecological validity than speech recorded in laboratories and 2) male informants indicated certain situations that are prone to speak straighter, indeed (e.g., when meeting unfamiliar people or masculine-acting men; Panfili, 2011).

In order to investigate effects of and on sexual orientation, a group comparison is necessary. In contrast to studies comparing non-straight vs. straight classifications (e.g., Pierrehumbert et al., 2004; Munson et al., 2006; Rendall et al., 2008), the present research investigated self-ratings of and ratings by others as lesbian/gay vs. straight. Especially when assuming effects of sexual orientation on speech, future studies are recommended to treat sexual orientation in a differentiated manner allowing for representing intra-group variability (e.g., applying the 7-point Kinsey-like scale) and avoiding a coarse grouping. Connectedly, future studies addressing contextual effects related to sexual orientation on speech (Podesva, 2006) are recommended to rely on a comparison of at least two sexual orientation groups that are represented by different speakers.

Although the present research contributes to a more differentiated perspective on sexual orientation and gender, it limited sexual orientation and gender to participants who self-ascribed as and were sexually attracted to women or men. This is in line with previous face and voice research showing shortcomings in addressing more genders in the context of sexual orientation. However, to the best of my knowledge, the present research was the first one explicitly dealing with bisexual speakers (Manuscript 4; for face research see Ding & Rule, 2012). For a more comprehensive understanding of sexual orientation signals, future studies need to investigate individuals whose gender identities and sexual orientations transgress gender binarities (intersexual, transsexual, transgender, pansexual, asexual, queer, non-binary etc.).

In contrast to suggestions by social identity theorists (Tajfel & Turner, 1979; Hornsey, 2008), in the present research speakers' group membership was not made salient before recordings. Hence, the present research missed the opportunity to provoke speaking patterns in line with relevant social identity aspects. This could explain the few acoustic differences found for lesbian/gay vs. straight speakers in the present studies. Hence, future research is advised to make group membership salient before recording speech materials (e.g., by filling out the 7-point Kinsey-like scale).

In general, studies dealing with sexual minorities have at least one shortcoming in common: The generalizability of their evidence is limited (Sandfort, 1997; Dekker & Schäfer, 1999; Meyer & Wilson, 2009). This is mainly due to two reasons. First, there is no knowledge about the distribution of characteristics in sexual minority populations and it is difficult to achieve because of participants' self-selection tendency that biases the results. For example, those individuals who are uncomfortable with their sexual orientation are less likely to participate. In addition, apart from a certain level of overlap, different studies use different sexual orientation definitions and measurements as has been stated above. For example, the classification strategies in the present research differ for female (Manuscript 2) and male speakers (Manuscript 3) which limited the comparability of findings.

In particular, there is one shortcoming referring to sampling in sociophonetic studies on sexual orientation: Caused by mostly small samples, their test-power is low (see Manuscript 3). Although the present research was based on relatively large-scale samples, only large effects could have been detected. Hence, it is the responsibility of future sociophonetic research on sexual orientation to negotiate the most appropriate definition of sexual orientation, to reach an agreement on a consistent sexual orientation measurement, and to use large-scale samples in order to detect small to moderate effects.

The present thesis provides evidence for German only. Hence, it is unclear whether and how gender-role self-concept is acoustically indexicalized in other languages and cultures and if EPSOM's indirect route is confirmed by other countries. Moreover, analyzed acoustic parameters in the present thesis mainly refer to f0 and acoustic vowel space characteristics (except for Manuscript 3). Acoustic correlates of sexual orientation and gender-role self-concept need to be investigated more broadly, especially for female speakers. Moreover, nasometry applied for male speakers needs to be refined. In the present research, the nasalance index based on multiple speech materials instead of single sound surroundings. Possibly, nasalization occurs in certain sound contexts only. Hence, future sociophonetic research is recommend to analyze a larger set of acoustic parameters (e.g., voice onset time in utterance-final positions) and to provide evidence on the markers of gender-role self-concept for other languages.

3.4 Conclusion

The present research provides evidence that speech stereotypes with respect to women's and men's sexual orientation in Germany are inaccurate – especially regarding their implied homogeneity and stereotypic content – and it suggests that they can be explained as mere exaggerations of tiniest kernels of truth. Similarly, by emphasizing variability within sexual orientation groups (with particular regard to gender-role self-concept and its acoustic correlates) and by considering certain switching points (psychological target characteristics, implicit signals, and impressions related to sexual orientation), the Expression and Perception of Sexual Orientation Model (EPSOM) explains how sexual orientation information is transmitted from one individual to another and why sexual orientation cannot be perfectly discerned. Hence, the present research expands our knowledge on social markers of sexual orientation and gender, promotes the erosion of sexual orientation stereotypes, and bears the potential to reduce homonegative discrimination.

References

Abele, A. E. (2003). The dynamics of masculine-agentic and feminine-communal traits: Findings from a prospective study. *Journal of Personality and Social Psychology,* 85, 768–776. DOI: 10.1037/0022-3514.85.4.768.

Allport, G. W. (1954). *The nature of prejudice*. Massachusetts: Addison-Wesley.

Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 201–271). San Diego, CA: Academic Press.

American Psychological Association (2012). *Publication manual of the American Psychological Association.* Washington: American Psychological Association.

Andrews, M. L., & Schmidt, C. P. (1997). Gender presentation: Perceptual and acoustical analyses of voice. *Journal of Voice,* 11, 307–313.

Avery, J. D., & Liss, J. M. (1996). Acoustic characteristics of less-masculine-sounding male speech. *Journal of the Acoustical Society of America,* 99, 3738–3748.

Baeck, H., Corthals, P., & van Borsel, J. (2011). Pitch characteristics of homosexual males. *Journal of Voice,* 25, 211–214. DOI: 10.1016/j.jvoice.2010.10.019.

Bailey, J. M., & Zucker, K. J. (1995). Childhood sex-typed behavior and sexual orientation. *Developmental Psychology,* 31, 43–55. DOI:10.1037/0012-1649.31.1.43

Beere, C. A. (1990). *Gender roles: A handbook of tests and measures.* Westport, CT: Greenwood Press.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology,* 42, 155–162. DOI: 10.1037/h0036215

Biemans, M. A. J. (2000). *Gender variation in voice quality* (Doctoral dissertation, Catholic University Nijmegen, Nijmegen, Netherlands). Retrieved from http://www.lotpublications.nl/Documents/38_fulltext.pdf

Bodenhausen, G. V., Kang, S. K., & Peery, D. (2012). Social categorization and the perception of social groups. In S. T. Fiske & C. N. Macrae (Eds.), *SAGE handbook of social cognition* (pp. 311-329). Los Angeles, CA: Sage. Retrieved from http://faculty.wcas.northwestern.edu/bodenhausen/BKP2012.pdf

Bruckmüller, S. (2013). Singled out as the effect to be explained: Implications for collective self-esteem. *Personality and Social Psychology Bulletin,* 39, 237–249. DOI: 10.1177/0146167212471686.

Byrd, D. (1992). Preliminary results on speaker-dependent variation in the TIMIT database. *Journal of the Acoustical Society of America*, 92, 593–596.

Cameron, D. (2005). Language, gender, and sexuality: Current issues and new directions. *Applied Linguistics,* 26, 482–502. DOI: 10.1093/applin/ami027.

Camp, M. (2009). *Japanese lesbian speech: Sexuality, gender identity, and language*. (Doctoral dissertation, University of Arizona, Columbus, OH). Retrieved from http://arizona.openrepository.com/arizona/bitstream/10150/195371/1/azu_etd_10574_sip1_m.pdf

Cartei, V., & Reby, D. (2012). Acting gay: Male actors shift the frequency components of their voices towards female values when playing homosexual characters. *Journal of Nonverbal Behavior,* 36, 79–93. DOI: 10.1007/s10919-011-0123-4.

Chiang, S. W. W. (2003). *Sounding 'gay': Prosodic patterns in gay speech in Singapore.* (Master thesis, National University of Singapore, Singapore) Retrieved from http://scholarbank.nus.edu.sg/handle/10635/14832

Coleman, R. O. (1971). Male and female voice quality and its relationship to vowel formant frequencies. *Journal of Speech and Hearing Research,* 14, 565–577.

Cox, W. T. L., Devine, P. G., Bischmann, A. A., & Hyde, J. S. (2015). Inferences about sexual orientation: The roles of stereotypes, faces, and the gaydar myth. *The Journal of Sex Research,* 53, 157–171. DOI: 10.1080/00224499.2015.1015714.

Crist, S. (1997). Duration of onset consonants in gay male stereotyped speech. *University of Pennsylvania Working Papers in Linguistics,* 4. Retrieved from http://repository.upenn.edu/cgi/viewcontent.cgi?article=1375&context=pwpl

Dekker, A., & Schäfer, S. (1999). Zum Sampling-Bias empirischer Studien über männliche Homosexualität [On the sampling-bias in empirical studies on male homosexuality]. *Zeitschrift für Sexualforschung,* 12, 350–361.

Diekman, A. B., & Eagly, A. H. (2000). Stereotypes as dynamic constructs: Women and men of the past, present, and future. *Personality and Social Psychology Bulletin,* 26, 1171–1188. DOI: 10.1177/0146167200262001

Ding, J. Y. C., & Rule, N. O. (2012). Gay, straight, or somewhere in between: Accuracy and bias in the perception of bisexual faces. *Journal of Nonverbal Behavior,* 36, 165–176.

Dunkle, J. H., & Francis, P. L. (1990). The role of facial masculinity/femininity in the attribution of homosexuality. *Sex Roles,* 23, 157–167.

Ebert, I. D., Steffens, M. C., & Kroth, A. (2014). Warm, but maybe not so competent? – Contemporary implicit stereotypes of women and men in Germany. *Sex Roles,* 70, 359–375. DOI:10.1007/s11199-014-0369-5

Eckert, P., & Podesva, R. J. (2011). Sociophonetics and sexuality: Towards a symbiosis of sociolinguistics and laboratory phonology. *American Speech,* 86, 6–13.

Ericsdotter, C., & Ericsson, A. M. (2001). Gender differences in vowel duration in read Swedish: Preliminary results. *Lund Working Papers*, 49, 1–4.

Evers, A., & Sieverding, M. (2014). Why do highly qualified women (still) earn less? Gender differences in long-term predictors of career success. *Psychology of Women Quarterly,* 38, 93–106. DOI: 10.1177/0361684313498071.

Fant, G. (1966). A note on vocal tract size factors and non-uniform f-pattern scalings. Quarterly *Progress and Status Report, Computational Linguistics*, 29, 22–30. Retrieved from http://www.speech.kth.se/prod/publications/files/qpsr/1966/1966_7_4_022-030.pdf

Fasoli, F., Maass, A., Paladino, M. P., & Sulpizio, S. (2017). Gay- and lesbian-sounding auditory cues elicit stereotyping and discrimination. *Archives of Sexual Behavior*, 46, 1261-1277. DOI 10.1007/s10508-017-0962-0

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology,* 23, 1–74.

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics,* 34, 409–438.

Freed, A. F. (1995). Language and gender. *Annual Review of Applied Linguistics,* 15, 3–22.

Freeman, J. B., Johnson, K. L., Ambady, N., & Rule, N. O. (2010). Sexual orientation perception involves gendered facial cues. *Personality and Social Psychology Bulletin,* 36, 1318–1331.

Gaudio, R. P. (1994). Sounding gay: Pitch properties in the speech of gay and straight men. *American Speech,* 69, 30–57.

Günzburger, D. (1989). Voice adaptation by transsexuals. *Clinical Linguistics & Phonetics,* 3, 163–172.

Guzik, K. M. (2004). Acoustic analysis of phonetic parameters of less masculine sounding German speech. In J. Harrington & C. Mooshammer (Eds.), *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK),* 36, 15–29. Retrieved from http://www.personal.uni-jena.de/~x1siad/papers/aipuk33.pdf

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis. A regression-based approach*. New York: Guilford Press.

Henton, C. G. (1989). Fact and fiction in the description of female and male pitch. *Language & Communication*, 9, 299–311.

Henton, C. G., & Bladon, R. A. W. (1985). Breathiness in normal female speech: Inefficiency versus desirability. *Language and Communication*, 5, 221–227.

Henton, C. G., & Bladon, R. A. W. (1988). Creak as a sociophonetic marker. In L. M. Hyman & C. N. Li (Eds.), *Language, speech and mind: Studies in honour of Victoria A. Fromkin*, (pp. 3–29). London: Routledge.

Herek, G. M., Cogan, J. C., & Gillis, J. R. (2002). Victim experiences in hate crimes based on sexual orientation. *Journal of Social Issues*, 58, 319–339.

Hornsey, M. J. (2008). Social identity theory and self-categorization theory: A historical review. *Social and Personality Psychology Compass*, 2, 204–222. DOI: 10.1111/j.1751-9004.2007.00066.x

Hughes, S. M., & Bremme, R. (2011). The effects of facial symmetry and sexually-dimorphic facial proportions on assessments of sexual orientation. *Journal of Social, Evolutionary, and Cultural Psychology,* 5, 214–230.

Jacobs, G. (1996). Lesbian and gay male language use: A critical review of the literature. *American Speech,* 71, 49–71.

Jacobs, G., Smyth, R., & Rogers, H. (2000). Language and sexuality: Searching for the phonetic correlates of gay- and straight-sounding male voices. *Toronto Working Papers in Linguistics*, 18, 46–61.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, 108, 1252–1263

Judd, C. M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review,* 100, 109–128. DOI: 10.1037/0033-295X.100.1.109

Jussim, L. (2012). *Social perception and social reality: Why accuracy dominates bias and self-fulfilling prophecy*. New York, NY: Oxford University Press.

Jussim, L., Crawford, J. T., & Rubinstein, R. S. (2015). Stereotype (in)accuracy in perceptions of groups and individuals. *Current Directions in Psychological Science*, 24, 490–497.

Kachel, S., Simpson, A.P., & Steffens, M.C. (2017). Acoustic correlates of sexual orientation and gender-role self-concept in women's speech. *Journal of the Acoustical Society of America*, 141, 4793–2809. DOI: 10.1121/1.4988684

Kachel, S., Steffens, M.C., & Niedlich, C. (2016). Traditional masculinity and femininity: Validation of a new scale assessing gender roles. *Frontiers in Personality and Social Psychology*, 7:956. DOI: 10.3389/fpsyg.2016.00956

Katz-Wise, S. L., & Hyde, J. S. (2012). Victimization experiences of lesbian, gay, and bisexual individuals: A meta-analysis. *Journal of Sex Research,* 49, 142–167. DOI: 10.1080/00224499.2011.637247.

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008). Tandem-straight: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, f0, and aperiodicity estimation. *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2008*, 3933–3936.

Kawahara, H., Takahashi, T., Morise, M., & Banno, H. (2009). Development of exploratory research tools based on tandem-straight. *Proceedings: APSIPA ASC 2009: Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference*, 111–120.

Kinsey, A., Pomeroy, W., & Martin, C. (1948). *Sexual behavior in the human male*. Philadelphia: Saunders.

Kite, M. E., & Deaux, K. (1987). Gender belief systems: Homosexuality and the implicit inversion. *Psychology of Women Quarterly,* 11, 83–96. DOI: 10.1111/j.1471-6402.1987.tb00776.x.

Kirkshaw, S. (2015). Intersectionality and the social meanings of variation: Class, ethnicity, and social practice. *Language in Society*, 44, 629–652.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America,* 87, 820–857.

Kouroupetroglou, G., & Chrysochoidis, G. (2014). Formant tuning in Byzantine chanting. *First International Interdisciplinary Conference "The Psaltic Art as an Autonoumous Science"*, Volos, Greece.

Lakoff, R. (1973). Language and woman's place. *Language in Society,* 2, 45–80.

Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., & Bourne, V. T. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. *Journal of the Acoustical Society of America*, 59, 675–678.

Lerman, J., & Damsté, P. (1969). Voice pitch of homosexuals. *Folia Phoniatrica,* 21, 340–346.

Linville, S. E. (1998). Acoustic correlates of perceived versus actual sexual orientation in men's speech. *Folia Phoniatrica et Logopaedica,* 50, 35–48.

Lippa, R. A. (2005). Sexual orientation and personality. *Annual Review of Sex Research,* 16, 119–153. DOI:10.1080/10532528.2005.10559831

Mack, S. (2010). Perception and identity: Stereotypes of speech and sexual orientation in Puerto Rican Spanish. In C. Borgonovo, M. Español-Echevarría, & P. Prévost (Eds.), *Selected Proceedings of the 12th Hispanic Linguistics Symposium* (pp. 136-147). Cascadilla Proceedings Project. Retrieved from http://www.lingref.com/cpp/hls/12/paper2412.pdf

Mack, S., & Munson, B. (2012). The influence of /s/ quality on ratings of men's sexual orientation: Explicit and implicit measures of the 'gay lisp' stereotype. *Journal of Phonetics,* 40, 198–212. DOI: 10.1016/j.wocn.2011.10.002.

Madon, S. (1997). What do people believe about gay males? A study of stereotype content and strength. *Sex Roles,* 37, 663–685.

Mattingly, I. G. (1966). Speaker variation and vocal tract size. *Journal of the Acoustical Society of America,* 39, 1219.

McConnell-Ginet, S. (1983). Intonation in a man's world. In B. Thorne, C. Kramarae, & N. Henley (Eds.), *Language, Gender and Society* (pp. 69–88). Rowley: Newbury.

Meyer, I. H., & Wilson, P. A. (2009). Sampling lesbian, gay, and bisexual populations. *Journal of Counseling Psychology,* 56, 23–31. DOI: 10.1037/a0014587.

Mishima, K., Sugii, A., Yamada, T., Imura, H., & Sugahara, T. (2008). Dialectal and gender differences in nasalance scores in a Japanese population. *Journal of Cranio-Maxillofacial Surgery*, 36, 8–10.

Moonwomon-Baird, B. (1997). Toward a study of lesbian speech. In A. Livia & K. Hall (Eds.), *Queerly phrased. Language, Gender, and Sexuality* (pp. 202–213). Oxford: Oxford University Press.

Munson, B. (2007). The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation. *Language and Speech,* 50, 125–142. DOI: 10.1177/00238309070500010601.

Munson, B. (2010). The Influence of Actual and Perceived Sexual Orientation on Diadochokinetic Rate in Women and Men. In International Speech Communication

Association (Ed.), *11th Annual Conference of the International Speech Communication Association* (pp. 1525–1528). Makuhari, Japan: Curran Associates.

Munson, B. (2011). Lavender lessons learned: Or, what sexuality can teach us about phonetic variation. *American Speech,* 86, 14–31.

Munson, B., & Babel, M. (2007). Loose lips and silver tongues, or, projecting sexual orientation through speech. *Language and Linguistics Compass,* 1, 416–449. DOI: 10.1111/j.1749-818x.2007.00028.x.

Munson, B., McDonald, E. C., DeBoe, N. L., & White, A. R. (2006). The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *Journal of Phonetics,* 34, 202–240. DOI: 10.1016/j.wocn.2005.05.003.

Ochs, E. (1991). Indexing gender. In A. Duranti & C. Goodwin (Eds.), *Rethinking context: Language as an Interactive Phenomenon* (pp. 336–358). Cambridge: Cambridge University Press.

Panfili, L. (2011). Vowel duration and perceptions of the gay accent. *Linguistics Honors Projects,* Paper 7. Retrieved from http://digitalcommons.macalester.edu/ling_honors/7/

Pearcey, S. M., Docherty, K. J., & Dabbs, J. M., jr. (1996). Testosterone and sex role identification in lesbian couples. *Physiology & Behavior,* 60, 1033–1035.

Pernet, C. R., & Belin, P. (2012). The role of pitch and timbre in voice gender categorization. *Frontiers in Psychology*, 3, 1–11.

Piccolo, F. (2008). *Listener-identified phonetic correlates of gay-, lesbian- and straight-sounding speech* (Doctoral dissertation, University of Hawaii). Retrieved from https://scholarspace.manoa.hawaii.edu/handle/10125/20680

Pierrehumbert, J. B., Bent, T., Munson, B., Bradlow, A. R., & Bailey, M. J. (2004). The influence of sexual orientation on vowel production (L). *Journal of the Acoustical Society of America*, 116, 1905–1908.

Plöderl, M. (2014). Bayesian advice for gaydar-based picking up: Commentary on Lyons, Lynch, Brewer, and Bruno (2013). *Archives of Sexual Behavior,* 43, 7–9.

Podesva, R. J. (2006). Intonational variation and social meaning: Categorical and phonetic aspects. *University of Pennsylvania Working Papers in Linguistics*, 12, 189–202.

Podesva, R. J. (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics,* 11, 478–504.

Podesva, R. J. (2011). The California vowel shift and gay identity. *American Speech,* 86, 32–51.

Prothro, E. T., & Melikian, L. H. (1955). Studies in stereotypes: V. Familiarity and the kernel of truth hypothesis. *Journal of Social Psychology,* 41, 3–10. DOI: 10.1080/00224545.1955.9714248

Queen, R. M. (1997). "I don't speak spritch". Locating lesbian language. In A. Livia & K. Hall (Eds.), *Queerly phrased. Language, Gender, and Sexuality* (pp. 233–256). Oxford: Oxford University Press.

Rakić, T., Steffens, M. C., & Mummendey, A. (2011). Blinded by the accent! The minor role of looks in ethnic categorization. *Journal of Personality and Social Psychology,* 100, 16–29. DOI: 10.1037/a0021522.

Renn, P. (2003). *Subtypes of male gayity: Speech, male sexual orientation, and childhood gender nonconformity* (Bachelor Thesis, University of Texas). Retrieved from http://homepage.psy.utexas.edu/homepage/class/Psy158H/PrevHonors/Z111/project.htm

Rendall, D., Vasey, P. L., & McKenzie, J. (2008). The queen's English: An alternative, biosocial hypothesis for the distinctive features of "gay speech". *Archives of Sexual Behavior*, 37, 188–204.

Rieger, G., Linsenmeier, J. A. W., Gygax, L., & Bailey, J. M. (2008). Sexual orientation and childhood gender nonconformity: Evidence from home videos. *Developmental Psychology,* 44, 46–58.

Rieger, G., Linsenmeier, J. A. W., Gygax, L. Garcia, S., & Bailey, J. M. (2010). Dissecting "gaydar": Accuracy and the role of masculinity–femininity. *Archives of Sexual Behavior,* 39, 124–140. DOI: 10.1007/s10508-008-9405-2.

Robb, M., Gilbert, H., & Lerman, J. (2005). Influence of gender and environmental setting on voice onset time. *Folia Phoniatrica et Logopaedia*, 57, 125–133.

Rule, N. O., & Ambady, N. (2008). Brief exposures: Male sexual orientation is accurately perceived at 50ms. *Journal of Experimental Social Psychology,* 44, 1100–1105. DOI: 10.1016/j.jesp.2007.12.001.

Rule, N. O., Ambady, N., Adams, R. B., & Macrae, C. N. (2008). Accuracy and awareness in the perception and categorization of male sexual orientation. *Journal of Personality and Social Psychology,* 95, 1019–1028. DOI: 10.1037/a0013194.

Rule, N. O., Ambady, N., & Hallett, K. C. (2009). Female sexual orientation is perceived accurately, rapidly, and automatically from the face and its features. *Journal of Experimental Social Psychology,* 45, 1245–1251. DOI: 10.1016/j.jesp.2009.07.010.

Runge, T. E., Frey, D., Gollwitzer, P. E., Helmreich, R. L., & Spence, J. T. (1981). Masculine (instrumental) and feminine (expressive) traits: A comparison between students in the United States and West Germany. *Journal of Cross-Cultural Psychology,* 12, 142–162.

Russell, E. (2015). Sounding gay and sounding straight. The performance of male sexual identity in Italian. *Journal of Language and Sexuality,* 4, 30–76.

Russell, E. (2017). Style shifting and the phonetic performance of gay vs. straight. A case study from French. *Journal of Language and Sexuality,* 6, 128–176.

Ryan, C. S. (2002). Stereotype accuracy. *European Review of Social Psychology*, 13, 75–109.

Sandfort, T. G. M. (1997). Sampling male homosexuality. In J. Bancroft (Ed.), *Researching sexual behavior: Methodological issues* (pp. 261–275). Bloomington: Indiana University Press.

Sczesny, S., Bosak, J., Neff, D., & Schyns, B. (2004). Gender stereotypes and the attribution of leadership traits: A cross-cultural comparison. *Sex Roles,* 51, 631–645.

Simpson, A. P. (1998). Phonetische Datenbanken des Deutschen in der empirischen Sprachforschung und der phonologischen Theoriebildung [Phonetic databases of German in empirical linguistic research and phonological theory formation]. In K. J. Kohler (Ed.), *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK),* 33. Retrieved from http://www.personal.uni-jena.de/~x1siad/papers/aipuk33.pdf

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, 3, 621–640.

Simpson, A. P., & Ericsdotter, C. (2007). Sex-specific differences in f0 and vowel space. *Proceedings of the the 16th International Congress of Phonetic Sciences,* Saarbrücken, Germany., 933–936.

Singh, D., Vidaurri, M., Zambarano, R. J., & Dabbs, J. M., jr. (1999). Lesbian erotic role identification: Behavioral, morphological, and hormonal correlates. *Journal of Personality and Social Psychology,* 76, 1035–1049.

Sisson, C. (2003). *Phonetic cues in the evaluation of gay male speech in Canadian English and Québec French.* Pennsylvania State University. Retrieved from http://www.cs.cmu.edu/~./csisso/docs/SocioII-paper.pdf

Skorska, M. N., Geniole, S. N., Vrysen, B. M., McCormick, C. M., & Bogaert, A. F. (2015). Facial structure predicts sexual orientation in both men and women. *Archives of Sexual Behavior,* 44, 1377–94*.* DOI: 10.1007/s10508-014-0454-4.

Skuk, V. G. (2014). *The perception of speaker gender and speaker identity in the human voice* (Doctoral dissertation, Friedrich-Schiller-University Jena, Jena, Germany).

Smyth, R., Jacobs, G., & Rogers, H. (2003). Male voices and perceived sexual orientation: An experimental and theoretical approach. *Language in Society,* 32, 329–350.

Smyth, R., & Rogers, H. (2008). Do gay-sounding men speak like women? *Toronto Working Papers in Linguistics*, 27, 129–144.

Spence, J. T., & Buckner, C. E. (2000). Instrumental and expressive traits, trait stereotypes, and sexist attitudes. *Psychology of Women Quarterly,* 24, 44–62. DOI: 10.1111/j.1471-6402.2000.tb01021.x

Steffens, M.C., Niedlich, C., Kachel, S., & Methner, N. (2016). Impression formation of applicants differing in sexual orientation: An attempt to integrate theoretical models and a review of the empirical evidence. In F. Earley (Ed.), *Sexual orientation: Perceptions, discrimination and acceptance.* New York: Nova Science Publishers.

Stern, C., West, T. V., Jost, J. T., & Rule, N. O. (2013). The politics of gaydar: Ideological differences in the use of gendered cues in categorizing sexual orientation. *Journal of Personality and Social Psychology,* 104, 520–541.

Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology,* 18, 86–100.

Stuart-Smith, J. (2007). Empirical evidence for gendered speech production: /s/ in Glaswegian. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology* (pp. 65–86). New York: Mouton de Gruyter.

Sulpizio, S., Fasoli, F., Maass, A., Paladino, M. P., Vespignani, F., Eyssel, F., & Bentler, D. (2015). The sound of voice: Voice-based categorization of speakers' sexual orientation within and across languages. *PLoS ONE,* 10, e0128882. DOI: 10.1371/journal.pone.0128882

Tabak, J. A., & Zayas, V. (2012). The roles of featural and configural face processing in snap judgments of sexual orientation. *PLoS ONE,* 7, e36671. DOI:10.1371/journal.pone.0036671

Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *Psychology of* Intergroup *Relations* (pp. 33–47). Monterey, Calif: Brooks/Cole Pub. Co.

Terango, L. (1966). Pitch and duration characteristics of oral reading. *Journal of Speech & Hearing Research,* 9, 590–595.

Trepte, S. (2006). Social identity theory. In J. Bryant & P. Vorderer (Eds.), *Psychology of Entertainment* (pp.255–272). New Jersey: Lawrence Erlbaum Associates.

Tskhay, K. O., Feriozzo, M. M., & Rule, N. O. (2013). Facial features influence the categorization of female sexual orientation. *Perception,* 42, 1090–1094.

Tskhay, K. O., & Rule, N. O. (2013). Accuracy in categorizing perceptually ambiguous groups: A review and meta-analysis. *Personality and Social Psychology Review,* 17, 72–86. DOI: 10.1177/1088868312461308.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group. A self-categorization theory.* New York, NY: Basil Blackwell.

Valentova, J. V., & Havlíček, J. (2013). Perceived sexual orientation based on vocal and facial stimuli is linked to self-rated sexual orientation in Czech men. *PLoS ONE,* 8, e82417. DOI: 10.1371/journal.pone.0082417.

Valentova, J. V., Kleisner, K., Havlíček, J., & Neustupa, J. (2014). Shape differences between the faces of homosexual and heterosexual men. *Archives of Sexual Behavior,* 43, 353–361. DOI: 10.1007/s10508-013-0194-x.

van Borsel, J., Vandaele, J., & Corthals, P. (2013). Pitch and pitch variation in lesbian women. *Journal of Voice,* 27, 656.e13. DOI: 10.1016/j.jvoice.2013.04.008.

van Lierde, K. M., Wuyts, F. L., De Bodt, M., & Van Cauwenberge, P. (2001). Nasometeric values for normal nasal resonance in the speech of young Flemish adults. *Cleft Palate-Craniofacial Journal,* 38, 112–118.

Waksler, R. (2001). Pitch range and women's sexual orientation. *WORD,* 52, 69–77.

Warner, M. (1993). *Fear of a Queer Planet.* University of Minnesota Press, Minneapolis.

Webbink, P. (1981). Nonverbal behavior and lesbian/gay orientation. In C. Mayo & N. M. Henley (Eds.), *Gender and Nonverbal Behavior* (pp. 253–259). New York, NY: Springer.

Whiteside, S. P., & Irving, K. J. (1997). Speakers' sex differences in voice onset time: Some preliminary findings. *Perceptual and Motor Skills,* 85, 459–63.

Wilde, A., & Diekman, A. B. (2005). Cross-cultural similarities and differences in dynamic stereotypes: A comparison between Germany and the United States. *Psychology of Women Quarterly,* 29, 188–196. DOI:10.1111/j.1471-6402.2005.00181.x

Zimman, L. (2010). Female-to-male transsexuals and gay-sounding voices: A pilot study. *Colorado Research in Linguistics*, 22, 1–21.

Zwicky, A. M. (1997). Two lavender issues for linguists. In A. Livia & K. Hall (Eds.), *Queerly phrased. Language, gender, and sexuality* (pp. 21–34). New York: Oxford University Press.

Appendix

List of Manuscripts

Manuscript 1. Kachel, S., Steffens, M. C., & Niedlich, C. (2016). Traditional Masculinity and Femininity: Validation of a New Scale Assessing Gender Roles. *Frontiers in Psychology,* 7:956. DOI: 10.3389/fpsyg.2016.00956.

Manuscript 2. Kachel, S., Simpson, A. P., & Steffens, M. C. (2017). Acoustic Correlates of Sexual Orientation and Gender-Role Self-Concept in Women's Speech. *Journal of the Acoustical Society of America,* 141, 4793- 4809. DOI: 10.1121/1.4988684

Manuscript 3. Kachel, S., Simpson, A. P., & Steffens, M. C. (2017). Do I Sound Straight? – Acoustic Correlates of Actual and Perceived Sexual Orientation and Masculinity/Femininity in Men's Speech. In press in *Journal of Speech, Language, and Hearing Research.*

Manuscript 4. Kachel, S., Radtke, A., Skuk, V. G., Zäske, R., Simpson, A. P., & Steffens, M. C. (2017). Do They All Speak the Same? – Investigating Sexual Orientation Information Using Voice Averages. Invited to Revise and Resubmit in *PLOS ONE*.

Manuscript 5. Kachel, S., Steffens, M. C., & Simpson, A. P. (2017). The Expression and Perception of Sexual Orientation Model: Speech Based Evidence. Submitted to *Journal of Personality and Social Psychology.*

Manuscript 1

Kachel, S., Steffens, M. C., & Niedlich, C. (2016).

Traditional Masculinity and Femininity: Validation of a New Scale Assessing Gender Roles.

Frontiers in Psychology, 7:956. DOI: 10.3389/fpsyg.2016.00956.

frontiers in Psychology

ORIGINAL RESEARCH
published: 05 July 2016
doi: 10.3389/fpsyg.2016.00956

Traditional Masculinity and Femininity: Validation of a New Scale Assessing Gender Roles

Sven Kachel, Melanie C. Steffens* and Claudia Niedlich*

Department of Social and Economic Psychology, University of Koblenz and Landau, Landau, Germany

Gender stereotype theory suggests that men are generally perceived as more masculine than women, whereas women are generally perceived as more feminine than men. Several scales have been developed to measure fundamental aspects of gender stereotypes (e.g., agency and communion, competence and warmth, or instrumentality and expressivity). Although omitted in later version, Bem's original Sex Role Inventory included the items "masculine" and "feminine" in addition to more specific gender-stereotypical attributes. We argue that it is useful to be able to measure these two core concepts in a reliable, valid, and parsimonious way. We introduce a new and brief scale, the Traditional Masculinity-Femininity (TMF) scale, designed to assess central facets of self-ascribed masculinity-femininity. Studies 1–2 used known-groups approaches (participants differing in gender and sexual orientation) to validate the scale and provide evidence of its convergent validity. As expected the TMF reliably measured a one-dimensional masculinity-femininity construct. Moreover, the TMF correlated moderately with other gender-related measures. Demonstrating incremental validity, the TMF predicted gender and sexual orientation in a superior way than established adjective-based measures. Furthermore, the TMF was connected to criterion characteristics, such as judgments as straight by laypersons for the whole sample, voice pitch characteristics for the female subsample, and contact to gay men for the male subsample, and outperformed other gender-related scales. Taken together, as long as gender differences continue to exist, we suggest that the TMF provides a valuable methodological addition for research into gender stereotypes.

Keywords: gender stereotypes, gender roles, gender-role self-concept, femininity, masculinity, actual and perceived sexual orientation, scale construction, voice pitch characteristics

OPEN ACCESS

Edited by:
Marcel Zentner, University of Innsbruck, Austria

Reviewed by:
Laura Mezquita, Universitat Jaume I, Spain
Ulrich Klocke, Humboldt-Universität zu Berlin, Germany

****Correspondence:***
Sven Kachel
sven.kachel@uni-jena.de
Melanie C. Steffens
steffens@uni-landau.de

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *02 August 2015*
Accepted: *09 June 2016*
Published: *05 July 2016*

Citation:
Kachel S, Steffens MC and Niedlich C (2016) Traditional Masculinity and Femininity: Validation of a New Scale Assessing Gender Roles. Front. Psychol. 7:956. doi: 10.3389/fpsyg.2016.00956

Abbreviations: BSRI, Bem Sex Role Inventory; BSRI-F, Bem Sex Role Inventory-femininity scale; BSRI-M, Bem Sex Role Inventory-masculinity scale; CGRB, Childhood Gender Role Behavior; f0, fundamental frequency; GEPAQ, German Extended Personal Attributes Questionnaire; GEPAQ-F, German Extended Personal Attributes Questionnaire-femininity scale; GEPAQ-M, German Extended Personal Attributes Questionnaire-masculinity scale; GRB, Gender Role Behavior; GRB-F, Gender Role Behavior-femininity scale; GRB-M, Gender Role Behavior-masculinity scale; IAT, Implicit Association Test; PAQ, Personal Attributes Questionnaire; TMF, Traditional Masculinity-Femininity; TMF-F, Traditional Masculinity-Femininity-femininity scale; TMF-M, Traditional Masculinity-Femininity-masculinity scale.

INTRODUCTION

Every time a group of people is addressed as "Ladies and Gentlemen!" the pervasiveness of gender over all other social categories is demonstrated. Gender is also one of the first social categories that children learn in today's societies, and thus knowledge of gender stereotypes is evident from early childhood on (for a recent review, see Steffens and Viladot, 2015) and into adulthood, with both adolescents and college students construing their self-concepts in line with the gender stereotypes they have internalized (e.g., Nosek et al., 2002; Steffens et al., 2010). Since the 1970s, following Bem's (1974) pioneering work, many scales have been designed, developed, and widely used for measuring traits traditionally considered as typically male vs. typically female (Constantinople, 1973). In recent years, such measures have often failed to find between-gender differences in self-ascriptions of gender stereotypical traits (e.g., Sczesny et al., 2004), which is presumably due to changes in gender roles across the decades (e.g., Diekman and Eagly, 2000; Wilde and Diekman, 2005; Ebert et al., 2014). Still, gender differences in self-ascriptions do continue to exist, and there are attempts to measure different aspects of masculinity and femininity, including, for example, everyday behavior such as housework (Athenstaedt, 2003). In the present paper, we argue that a scale that reliably and validly measures differences in an individual's underlying conceptualization of his or her own masculinity-femininity would be valuable for gender research. To date, these constructs can only be measured using two items, "masculine" and "feminine," which is somewhat limited given that established standards of psychological assessment typically recommend using a larger number of items (e.g., Bühner, 2010). In the present article, we introduce a new, extended, but still parsimonious scale, the Traditional Masculinity-Femininity Scale, TMF, to fill this gap. Using a known-groups approach, we present two studies testing this measure's reliability as well as its incremental and criterion validity, and we provide evidence for its convergent validity.

We define "traditional masculinity" and "traditional femininity" as relatively enduring characteristics encompassing traits, appearances, interests, and behaviors that have traditionally been considered relatively more typical of women and men, respectively (adapting the definitions provided by Constantinople, 1973). It is important to note that the focus of the present paper is on gender-related self-assessment. Complementary research has investigated many different aspects of gender, for example, gender-role norms (e.g., Athenstaedt, 2000; Thompson and Bennet, 2015; Klocke and Lamberty, unpublished manuscript).

In a seminal study on masculinity and femininity, Deaux and Lewis (1984) investigated the perceived relationship between gender and gender-related components, such as role behaviors (e.g., head of household vs. takes care of children), traits, occupations, and physical characteristics (e.g., tall, broad-shouldered vs. soft voice, graceful). The researchers showed that these components were interdependent, impacting on one another, as well as on perceived gender and sexual orientation. In other words, participants readily generalized from one component to the others. In addition, physical appearance played a particularly large role. Such findings indicate that gender stereotypes may be based on some sort of "core" masculinity and femininity. Similarly, individuals may use such "core" masculinity and femininity in their self-construal.

The first attempts to gauge masculinity and femininity placed these constructs on a bipolar spectrum and involved measuring simple collections of personality traits on which women and men differed on average (for a review, see Constantinople, 1973). By contrast, Bem's pioneering Sex Role Inventory (BSRI; Bem, 1974) used gender-stereotypical traits to independently measure masculinity and femininity (e.g., masculine items such as competitive and dominant, and feminine items such as affectionate and gentle). She pointed out that women/men who score high on both scales were called androgynous. Importantly, "masculine" and "feminine" were included as items in these original scales, but were excluded from the revised version (Bem, 1979) because of problematic loadings on the factors on which the masculine and feminine traits loaded, respectively. Exploratory factor analyses showed an instable factor structure but often converged on three-factor solutions: Masculine traits on one factor, feminine traits on a second factor, and masculine-feminine along with participant gender on a third factor (e.g., Niedlich et al., 2015, see review by Choi and Fuqua, 2003). It has thus been suggested that the two independent masculinity and femininity trait dimensions are complemented by one bipolar masculinity-femininity dimension (see Constantinople, 1973; Spence et al., 1975; Bem, 1979) that reflects gender identity instead of gender-role related aspects (e.g., Bem, 1979; Spence and Buckner, 2000). As Choi and Fuqua (2003) suggest, inventories such as the BSRI "may not capture the complex and multidimensional nature of masculinity/femininity." Instead, "masculinity and femininity could be two higher order constructs, with each having its own subconstructs" (p. 873). Similar to other scales (e.g., Personal Attributes Questionnaire, PAQ, by Spence et al., 1975), the BSRI appears to tap more specific constructs, often referred to as instrumentality/agency and expressivity/communion (e.g., Fiske et al., 2002; Abele and Wojciszke, 2007), rather than masculinity and femininity in general. For the present purposes it is important to note that if masculinity and femininity are directly measured they should load on one bipolar masculinity-femininity dimension.

Another limit to the practical use of these established scales pertains to the generally small magnitude of gender differences found on these two dimensions (e.g., Deaux, 1984). In other words, women and men appear rather similar on "masculinity" and "femininity." More recently, gender differences have not emerged at all between graduates with the same major (see Abele, 2000). In short, scales that have been developed to assess aspects of masculinity and femininity have recently failed to find gender differences (see also Sczesny et al., 2004; Evers and Sieverding, 2014). This could indicate that gender differences in masculinity and femininity are a thing of the past (Alvesson, 1998). However, it could also mean that the scales do not tap the most relevant aspects of the constructs on which gender differences continue to exist. For example, gender roles have changed over the last decades, particularly women's roles, so that today's women possess more of the traits traditionally considered as masculine (e.g., Diekman and Eagly, 2000; Spence and Buckner, 2000; Wilde and Diekman, 2005; Ebert et al., 2014). According to these findings, instrumental traits have become more socially desirable for women and expressive traits have become more socially desirable for men (Swazina et al., 2004).

In order to overcome limitations of the discussed scales, there have been attempts to measure other aspects of masculinity and femininity to account for the multiple dimensions they are reflected in, such as physical appearance, behaviors, attitudes, and interests (e.g., Spence and Buckner, 2000; Blashill and Powlishta, 2009). For example, Athenstaedt (2003) observed considerable gender differences in everyday behavior such as "putting flowers on the desk" (feminine) and "putting the meat on the barbeque" (masculine), strongly suggesting the continued importance of gender differences. Complementing these existing approaches, we suggest directly assessing the presumed higher-order constructs, namely masculinity and femininity. However, instead of using only these two items, we constructed a scale that can be tested empirically with regard to its reliability and validity.

Scale Construction

We introduce the TMF scale, an instrument for measuring gender-role self-concept. Appendix A1 in Supplementary Material shows all items, both English translations and original German wordings. Each item initially included in scale construction was selected based on theoretical considerations, as outlined in the following. We argue that we can measure the "core" of masculinity/femininity by referring to three central aspects, identified by Constantinople (1973), that we summarize using the term gender-role self-concept: Namely, gender-role adoption, gender-role preference, and gender-role identity. Constantinople (1973) defines *gender-role adoption* as the actual manifestation (i.e., how masculine-feminine a person considers her- or himself) and *gender-role preference* as the desired degree of masculinity-femininity (i.e., how masculine-feminine a person ideally would like to be). According to Kagan (1964), *gender-role identity* refers to a comparison of gender-related social norms and the gender-related characteristics of the individual (e.g., how a person actually looks compared to expected gender-typical appearances according to societal norms). Hence, for gender-role identity social comparisons as well as references to different gender-related aspects are emphasized (e.g., looks, behaviors etc.), whereas gender-role adoption and preference are based on non-relative, absolute statements. Following the former approach, we use TMF as a reference point. Based on dimensions identified as important in previous research, the TMF encompasses gender-role identity with regard to physical appearance, behavior, interests, and attitudes and beliefs (e.g., Deaux and Lewis, 1984; Athenstaedt, 2003). As mentioned, *physical appearance* was shown to play a particularly large role in implicating other components of gender stereotypes (Deaux and Lewis, 1984). Athenstaedt (2003) advocated the inclusion of gender-stereotypical *behaviors* in addition to traits, so this domain was included in the TMF as well. Lippa (2008) found that gender-related *interests* were highly relevant in discriminating women and men as well as lesbians/gay men from straight people. Additionally, his study showed that instrumental and expressive traits were outperformed by these gender-related interests in predicting participants' gender. Consequently, we included gender-related interests in the TMF (instead of gender-related traits). Finally, regarding *attitudes and beliefs*, gender differences have often been found, for example, with regard to attitudes toward minority groups (e.g., Sidanius et al., 1994; Kite and Whitley, 1996). We therefore also included self-assessment of attitudes and beliefs in the TMF.

One advantage of the TMF is that each of the mentioned scale dimensions is measured on a global level and not by various specific indicator items. Different from the instruments described above, which infer masculinity-femininity from the degree of affirmation of specific traits and behaviors, the TMF aims to directly assess masculinity-femininity. For example, "Traditionally, my behavior would be considered as…" 1 (*not at all masculine*) to 7 (*very masculine*). We consider it an asset of the scale that it is thus independent of specific stereotype content regarding masculinity and femininity that depend on culture and time (e.g., intelligent and ambitious as masculine, childlike and shy as feminine, see BSRI; in the General Discussion we discuss how far this global conception can also be considered a limitation). The TMF consists of six items only: One for gender-role adoption ("I consider myself as…"), one for gender-role preference ("Ideally, I would like to be…"), and four for gender-role identity ("Traditionally, my 1. interests, 2. attitudes and beliefs, 3. behavior, and 4. outer appearance would be considered as…") in order to measure an individual's gender-role self-concept in a parsimonious way. All of them have high face validity. Each item is to be independently rated in terms of femininity and masculinity. A 7-point-scale is used to gauge the extent to which the participant feels feminine or masculine, how feminine or masculine she or he ideally would like to be, and how feminine and masculine her or his appearance, interests, attitudes, and behavior would traditionally be seen. Construct validity is tested in the studies described below. The TMF was used with masculinity and femininity as two unipolar dimensions (Study 1: 1, *not at all masculine*, to 7, *very masculine*, and 1, *not at all feminine*, to 7, *very feminine*) vs. one bipolar dimension (pilot study, Study 2; 1, *very masculine*, to 7, *very feminine*) in order to check for dimensionality.

Overview of the Present Research

We validated the TMF in various ways. First, we conducted an item analysis and a factor analysis. As suggested by findings reported by Bem (1979), Constantinople (1973), and Spence et al. (1975; see Introductory Section), the TMF's items should load on one factor and tap a one-dimensional masculinity-femininity construct. Hence, we expected the TMF to measure a one-dimensional gender-role self-concept (Hypothesis 1).

Validation by Using the Known-Groups Approach

Based on the idea that gender differences are not a thing of the past, as indicated in the introduction, a valid masculinity and femininity scale should show these gender differences. Therefore, we expected men and women to differ considerably on self-ascriptions on the TMF, with men being more masculine and less feminine than women (Hypothesis 2).

Moreover, a valid masculinity and femininity scale should show differences between people differing in sexual orientation. The essence of gender stereotypes of straight women and men is that they conform to traditional gender roles (e.g., Kite and Deaux, 1987; Kite and Whitley, 1996; Madon, 1997; Blashill and Powlishta, 2009). Lay people expect straight women to be more feminine and less masculine than lesbians, and straight men to be more masculine and less feminine than gay men. Similarly, straight women's and men's self-ascriptions are, on average, more gender-typed than those of lesbians and gay men (see meta-analysis by Lippa, 2005). Bisexual women were found to score on masculinity-femininity in between lesbians and straight women (Lippa, 2005). Therefore, we used the known-groups approach as an established method for testing a scale's validity (e.g., Howitt and Cramer, 2008). We expected lesbians' self-ascriptions on the TMF to be less feminine and more masculine compared to straight women (Hypothesis 3a). Bisexual women should score in between (Hypothesis 3b). Additionally, we expected straight men's self-ascriptions to be more masculine and less feminine compared to gay men (Hypothesis 3c).

Because straight women and men conform to gender roles more than lesbians and gay men, comparing lesbians and gay men constituted a stricter test of the TMF. Consistent with Hypothesis 2 and gender self-stereotyping but contradictory to implicit gender inversion theory (Kite and Deaux, 1987; which we turn to in General Discussion), we hypothesized lesbians to be more feminine and less masculine than gay men (Hypothesis 4).

The idea that differences in "core" masculinity and femininity underlie differences in lesbians' and gay men's vs. straight women and men's self-ascriptions in gender typicality can formally be conceived as masculinity-femininity mediating the relationship between sexual orientation and responses on scales such as the BSRI (Hypothesis 5).

Validation by Implicit and Explicit Gender-Related Measures

A common critique of self-report measures is that they could reflect differences in social desirability more than "true" underlying differences in traits. Using implicit measures relying on response-time differences, such as an Implicit Association Test (IAT), may minimize this problem (Greenwald et al., 1998). Implicit measures are assumed to assess the impulsive system: Habitual, repeated, long-term associations between concepts (Strack and Deutsch, 2004), including self-related concepts (e.g., Steffens and Schulze-Koenig, 2006). We expected lesbians to describe themselves more masculine and less feminine than straight women (Hypothesis 6).

Adults' masculinity-femininity is related to (recalled) gender conformity during adolescence (e.g., Safir et al., 2003) and childhood (e.g., Lippa, 2008). Thus, gender-role instruments for assessing current traits and behaviors as well as recalled gender-typical behaviors, preferences, and interests during childhood were also suitable for testing convergent validity. We assumed all these characteristics to show moderate correlations with the TMF (Hypothesis 7).

Additionally, we expected the TMF to predict sexual orientation within one gender group better than other gender-related scales. We assumed the TMF to outperform other gender-related scales when predicting sexual orientation of women and men (Hypothesis 8).

Hypotheses Based on Criterion Validity

As indicated above, lay people use gender-typicality as an indicator for judging someone's sexual orientation (Rieger et al., 2010; Valentova et al., 2011). People self-reporting gender-typical characteristics are likely to be perceived as straight, whereas people who do not display such characteristics are more likely to be perceived as lesbian or gay on pictures, videos, and speech recordings. Hence, targets who are perceived as straight could be those who self-describe as gender-typical in masculinity-femininity ratings (Hypothesis 9).

Additionally, there is some evidence that voice pitch characteristics, also called fundamental frequency features, of lesbians and gay men are shifted toward what is typical for straight women and men. Generally, compared to straight women, straight men show voice pitches that are lower on average, in variability, and in range (e.g., Pierrehumbert et al., 2004; Munson and Babel, 2007). Average voice pitch has been found to be lower in straight compared to gay men (Baeck et al., 2011) and higher in straight women compared lesbians (Camp, 2009). Hence, we assumed gender-typical masculinity-femininity self-ratings to be reflected in gender-typical patterns of voice pitch characteristics (Hypothesis 10).

Furthermore, contact frequency of straight women and men with lesbians and gay men is linked to attitudes toward them (e.g., Swank and Raiz, 2010): A lower contact frequency is connected to more negative beliefs about lesbians and gay men. One belief about lesbians and gay men is that they transgress gender roles, on average (e.g., Kite and Whitley, 1996). It thus seems plausible that people who are more gender-typical themselves are those who have less contact to lesbians and gay men and hold more negative beliefs. Hence, we assumed gender-typical masculinity-femininity self-ratings to be connected to more current contact with straight women and men and less current contact with lesbians and gay men (Hypothesis 11).

Hypotheses Concerning Test-Retest Reliability and Predictive Validity

Finally, the TMF was expected to show at least moderate test-retest reliabilities given that people were re-invited after a 1-years period (Hypothesis 12). From a scale validation perspective, it is desirable to present analyses in which the predictor is truly assessed before the criterion. Therefore, we expected at least moderate predictive validity for other gender-related features at second measurement (Hypothesis 13).

PILOT STUDY

The pilot study had two aims. First, we tested the factor structure of the scale's version that contained six bipolar items. We assumed the TMF items to load on one factor (Hypothesis 1). Additionally, we wanted to determine the appropriateness of every single item by using an item analysis. Second, we assessed the scale's validity using a known-groups approach (Hypothesis 2).

Methods

At the end of an online survey that had a different purpose, participants filled in the 6-item version of the TMF (see Appendix in Supplementary Material) and indicated their gender (response options: male, female, both, none, no response). Overall 319 participants finished the study. Thirteen of them were excluded from further analysis because they described themselves as both male and female or neither or they did not disclose their gender. Data from 188 women and 118 men were used for analysis. Their age ranged from 18 to 41 ($M = 23.6$, $SD = 3.1$). They were students of different majors from different German universities (specifically, in Thuringia). Participants received no compensation for participation. Approval for all studies reported in this paper was obtained by the board of ethics (=human subjects committee) of the School of Humanities and Social Sciences at the Friedrich-Schiller-University of Jena. All studies were carried out in accordance with its recommendations, with written informed consent obtained from all participants in accordance with the Declaration of Helsinki.

Results

In order to check for one-dimensionality of the TMF, an exploratory principal axis factoring (PAF) was conducted. Sample adequacy was confirmed by a Kaiser-Meyer-Olkin (KMO) criterion of 0.87. All items were suitable for factor analysis as indicated by item-specific KMO values >0.79 and moderate to high commonalities (0.57–0.88). According to a graphical scree-plot analysis, a one-factor solution was confirmed. There was a steep decline of explained variance from factor one (77%) to factor two (10%). Each of the six items was represented well by the factor (factor loadings ranged from 0.75 to 0.94).

Reliability of the TMF was high (Cronbach's $\alpha = 0.94$). As indicated by the coefficients in **Table 1**, no items needed to be deleted to improve reliability. Item-specific homogeneity was high and ranged from 0.66 to 0.72 (see **Table 1**). Corrected item-total correlations ranged from 0.72 to 0.91, suggesting that each item represented the scale well. Moreover, item means ranged from 0.51 to 0.59. Accordingly, every item received almost equal masculinity and femininity ratings, indicating that averaged across the sample containing women and men, items received "androgynous" responses, as one would expect. When computing item "difficulties" separately for each gender group, findings pointed in the expected directions: "Difficulties" ranged from 0.18 to 0.35 for the male sample, indicating "masculine" responses, and from 0.60 to 0.85 for the female sample, indicating "feminine" responses.

We found the expected bimodal distribution of the TMF scores. Men and women differed significantly in terms of the scale mean, $M_{male} = 2.56$ ($SD = 0.80$), $M_{female} = 5.28$ ($SD = 0.76$), $t_{(304)} = -29.83$, $p < 0.001$, and on every item, all $ts_{(287)} > -10.41$, all $ps < 0.001$. With the exception of two outlier individuals, the overlap between men's and women's scores was very small (see **Figure 1**). According to Kolmogorov-Smirnov statistics, the TMF scores were normally distributed for men ($Z = 0.99$, $p = 0.28$) and women ($Z = 0.78$, $p = 0.58$). Predicting gender by the TMF scores in a logistic regression analysis was 97% accurate [$B = 4.43$, $SE = 0.69$, $\chi^2_{(1)} = 41.38$, $p < 0.001$; Nagelkerke's $R^2 = 0.92$; Model $\chi^2_{(1)} = 347.87$, $p < 0.001$].

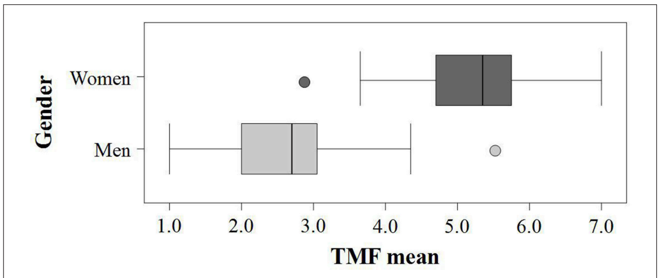


FIGURE 1 | Distribution of the TMF scores separately for men ($n = 118$) and women ($n = 188$) in the pilot study. The lines in the bars represent medians and bars indicate the range between 75th and 25th percentile. Error bars show the range of masculinity-femininity scores for non-outliers. Dots represent outlying values (1.5 *SD* above/below median).

Taken together, confirming Hypothesis 1, we found that the TMF tapped a one-dimensional construct which is in line with lay ascriptions and previous findings regarding the items masculine and feminine. All factor loadings were similar ($\Delta < 0.1$), so that an unweighted additive overall score was justified (Bortz and Döring, 2006). Its single items represented the overall scale very well and were strongly connected to each other. Hence, no item had to be excluded due to low item-specific homogeneity (Bortz and Döring, 2006). Moreover, confirming Hypothesis 2, the TMF was shown to discriminate between women and men at the scale and at the item level. Therefore, we kept all items in the TMF.

STUDY 1

The aim of Study 1 was to test the one-dimensionality, reliability, and validity of the TMF. We used a known-groups approach, with lesbians, bisexual, and straight women, to assess which of several gender-related scales is best in differentiating between these groups. In addition to the TMF, we used the BSRI as the gold standard in gender-related assessment. However, we also used the Gender Role Behavior Scale (GRB, Athenstaedt, 2003) and a newly created measure of childhood gender conformity (see Appendix in Supplementary Material). Moreover, an Implicit Association Test (IAT, Greenwald et al., 1998) was used to measure implicit associations of self with masculine vs. feminine.

We assumed that the TMF would reflect a one-dimensional masculinity-femininity construct (Hypothesis 1). Furthermore, we expected that on each measure, straight women would score higher on femininity and/or lower on masculinity as compared to lesbians (Hypothesis 3a). Bisexual women should score in between (Hypothesis 3b). Additionally, on an IAT (see below for details), we assumed straight women to associate more with feminine and less with masculine than lesbians (Hypothesis 6).

TABLE 1 | Item Characteristics of the TMF in the Pilot Study for the Whole Sample (left-hand values, $n = 306$) and Separately for Men (middle values, $n = 118$) and Women (right-hand values, $n = 188$).

	Corrected item-total correlation	Cronbach's α if item is deleted	Item means	Item homogeneity	Factor loading
1. I consider myself as …	0.91, 0.67, 0.63	0.91, 0.79, 0.76	0.56, 0.23, 0.77	0.78, 0.51, 0.46	0.94, 0.80, 0.76
2. Ideally, I would like to be…	0.87, 0.51, 0.56	0.92, 0.81, 0.77	0.55, 0.20, 0.76	0.76, 0.42, 0.42	0.91, 0.67, 0.74
3. Traditionally, my interests would be considered as…	0.77, 0.65, 0.56	0.93, 0.77, 0.77	0.51, 0.30, 0.63	0.69, 0.45, 0.41	0.84, 0.76, 0.71
4. Traditionally, my attitudes and beliefs would be considered as…	0.72, 0.61, 0.67	0.94, 0.80, 0.74	0.51, 0.35, 0.60	0.66, 0.45, 0.45	0.81, 0.74, 0.80
5. Traditionally, my behavior would be considered as…	0.85, 0.73, 0.67	0.93, 0.77, 0.74	0.52, 0.30, 0.66	0.75, 0.49, 0.45	0.90, 0.83, 0.79
6. Traditionally, my outer appearance would be considered as…	0.83, 0.45, 0.26	0.93, 0.82, 0.82	0.59, 0.18, 0.85	0.72, 0.35, 0.21	0.88, 0.61, 0.39

Scale ranged from 1—"very masculine" to 7—"very feminine."

Gender-related measures should be correlated with each other (Hypothesis 7), and scores on each measure should predict sexual orientation. We also tested the incremental validity of the TMF over the other measures. The TMF should predict sexual orientation better than other gender-related scales (Hypothesis 8). Finally, the TMF should measure a higher-order factor "core" masculinity-femininity that mediates effects of sexual orientation on other gender-related scales (Hypothesis 5). If women differ in masculinity-femininity based on their sexual orientation, indirect effects of the more specific masculinity-femininity related measures via the TMF on sexual orientation should be observed.

Method

Participants

Participants were 126 women from Germany and Luxembourg who took part in the study, voluntarily without compensation. Their age ranged from 19 to 47 years ($M = 31.13$, $SD = 8.52$). Participants were recruited either at the University of Trier or by a snowball technique. Given their scores on a Kinsey-like scale, they were divided into three groups of 47 straight women (Kinsey scores: 6–7), 32 bisexual women (3–5), and 47 lesbians (1–2). Most of the women were well educated, with 50% possessing university entrance qualifications and 40% holding a university degree. With $\alpha = 0.05$ and $N = 126$, based on Cohen's (1977) conventions, medium-size regression coefficients ($f^2 = 0.35$) could be detected with a statistical power of $1 - \beta = 0.95$ in a multiple linear regression with six predictors (Faul et al., 2007).

Materials

Implicit association test

In essence, IATs comprise two combined tasks in which stimuli that belong to four concepts are mapped onto two responses in different ways. IATs are based on the following idea: If someone is able to react relatively fast when two concepts share a response, these concepts appear to be associated for that person. In detail, stimuli were presented that represent the concepts *self, others, feminine,* and *masculine*. In one task, stimuli representing *self* or *feminine* required one response, and stimuli representing *others* or *masculine* required the other response (e.g., left vs. right key press). In the other task, stimuli representing *self* or *masculine* required one response, and stimuli representing *others* or *feminine* required the other response. A person considering herself feminine should be able to react faster in the self-feminine/others-masculine than in the self-masculine/others-feminine task.

We labeled one dimension for the IAT "typically feminine" vs. "typically masculine." The associated attributes presented were *feminine, female* vs. *masculine, male* (in German: *feminin, weiblich; maskulin, männlich*, see Steffens et al., 2008). The other dimension was "self" vs. "others." The stimuli on that dimension were synonyms of the superordinate concepts (*me, self* vs. *you, others*; in German: *Ich, Selbst; Du, Andere*). Participants were informed that concepts would be displayed throughout at the top left or right screen corner. Their task during the IAT would be to sort words belonging to these concepts by pressing the respective response key on the left or right as quickly as possible. A stimulus word would appear (e.g., feminine) after which participants would respond by pressing the appropriate key (e.g., left for *typically feminine*). The word would then be replaced by the next stimulus (e.g., *me*). Participants would again select the appropriate key (e.g., left for *self*). Each crucial, combined task consisted of four blocks of 62 trials. The order of the eight stimuli was randomized within each block, and the same eight stimuli were presented over and over. The reaction-stimulus interval was 200 ms. Missing reactions and errors led to an appropriate visual feedback (e.g., in case of errors, F! was shown for 200 ms). Participants received feedback on errors and reaction times after each block (e.g., given 10% errors or more: "You committed many errors. Please react more slowly and more correctly.").

The IAT effect was computed similar to the IAT D effect (Nosek et al., 2005, except that no "error penalty" was used, see Steffens et al., 2008): Specifically, the reaction time difference between the self-feminine/others-masculine and the self-masculine/others-feminine task was computed and divided by each individual's standard deviation across both tasks. In order to avoid artificially high scores obtained with very long scales, internal consistency was estimated based on the average reaction time difference in reaction to each of the eight stimuli. In other words, the IAT was treated as an eight item scale (following Steffens and Buchner, 2003). All internal consistencies are presented in **Table 2**.

Bem sex-role inventory

We translated the English short version of the BSRI (Bem, 1979) into German. It consisted of 30 items, 10 for the Masculinity Scale

TABLE 2 | Internal Consistencies (Cronbach's α, with number of items) and Correlations between Measures in Study 1.

	Alpha (items)	TMF –M	BSRI –F	BSRI –M	GRB –F	GRB –M	CGRB –F	CGRB –M	IAT effect
TMF-F	0.90 (6)	−0.85	0.42	(−0.07)	0.41	−0.51	0.71	−0.65	0.30
TMF-M	0.89 (6)		−0.30	(0.17)	−0.37	0.44	−0.60	0.57	−0.28
BSRI-F	0.83 (10)			(−0.08)	0.40	(−0.06)	0.21	(−0.11)	0.18
BSRI-M	0.78 (10)				(0.05)	(0.12)	(0.01)	(0.06)	−0.24
GRB-F	0.87 (29)					(−0.11)	0.40	−0.25	0.18
GRB-M	0.83 (23)						−0.48	0.47	(−0.10)
CGRB-F	0.88 (5)							−0.90	0.31
CGRB-M	0.88 (5)								−0.31
IAT effect	0.93 (8)								

All correlations are statistically significant at $\alpha \leq 0.05$ except for those in parentheses. Abbreviations: TMF, Traditional Masculinity-Femininity; BSRI, Bem Sex Role Inventory; GRB, Gender-Role Behavior; IAT effect: Differences in the implicit association test (IAT) between the self-masculine/others-feminine and the self-feminine/others-masculine task. Endings indicate masculinity (–M) und femininity (–F) scales. Higher scores indicate higher masculinity on masculinity scales and higher femininity on femininity scales.

(e.g., self-reliant, ambitious), 10 for the Femininity Scale (e.g., warm, tender), and 10 neutral items with a 7-point scale anchored 1 (*never applies*) to 7 (*always applies*). Participants were asked to rate the extent to which the given traits were adequate to describe them.

Traditional masculinity-femininity

The TMF was used as described in the Section Scale Construction with two unipolar dimensions, masculinity and femininity (12 items overall, see Appendix in Supplementary Material).

Childhood gender role behavior (CGRB)

Five items were used with a 7-point-scale in order to measure whether participants remembered to have been rather feminine during childhood, or rather typical girls, or not (see Appendix A2 in Supplementary Material). For example, we asked whether they had played with girls and girls' games, and whether they had liked wearing skirts and dresses.

Sexual orientation

As indicated in Section Participants, participants' sexual orientation was assessed using participants' responses on the item: "Regarding sexual orientation, I identify as …" (on a Kinsey-like scale, from 1 (*exclusively lesbian*) to 7 (*exclusively straight*). This was also the first item of a translated version of the Assessment of Sexual Orientation Scale (Coleman, 1987). Several additional items were originally used (sexual behavior: gender of partner and ideal partner; sexual fantasies, and emotional bindings). To be consistent with Study 2, we used only the first item to group participants as lesbians (scores 1–2), bisexual women (scores 3–5), and heterosexual women (scores: 6–7). The first item also correlated highly with the overall scale ($r = 0.95$), corroborating the decision to use only one item.

Gender role behavior scale

Participants rated themselves on a 7-point scale ranging from 1 (*not at all typical*) to 7 (*very typical*) on 52 everyday typically feminine or masculine behaviors (GRB, Athenstaedt, 2003; e.g., "watch soap operas," "change light bulbs").

Procedure

Participating students were tested at the University of Trier in a lab cubicle equipped with an iMac. The participants recruited via the snowball technique were tested individually in their homes or offices (as they wished) using an iBook. The instructions, the implicit tests, and the questionnaires were presented by a self-composed HyperCard computer program. Initially, participants were asked to report their age, educational background, and size of hometown. Then, they started with the IAT. IAT task order was held constant because of the correlational nature of the study (see e.g., Banse et al., 2001, for discussion). All participants did the self-masculine/others-feminine task first. After the IAT, the questionnaires were presented in the order described in the Materials Section—accordingly, data for the TMF was collected before all other scales. Finally, participants were debriefed and thanked.

Results

In all analyses in the present article, significance tests were conducted with $\alpha = 0.05$ and all statistical analyses were done with SPSS 22. One might suggest that all other scales in addition to the TMF used in the present research should also be submitted to factor analyses. However, commonalities of several of them were too low for conducting confirmatory factor analyses. To illustrate, in Study 2 we observed GRB-M (<0.01) and GRB-F (<0.10). Therefore, means of all established gender-related scales were computed according to the scales' theoretical basis as suggested by their authors.

Factor Analysis

In order to check for one-dimensionality of the TMF, an exploratory PAF with oblique rotation (oblimin: 0) was conducted for all 12 items. Sample adequacy was confirmed by a KMO criterion of 0.86. All items were suitable for factor analysis as indicated by item-specific KMO values >0.77 and moderate to high commonalities (0.50–0.80). Several indicators are in line with the same one-factor solution as in the Pilot Study and in Study 2 below. According to a graphical scree-plot analysis, a one-factor solution was confirmed. There was a steep decline of explained variance from factor one (61%) to factor two (12%). Moreover, the factor matrix showed a strong first factor suggesting all items to measure something similar.

An alternative confirmatory factor analysis with one factor replicating the findings of the Pilot Study yielded an overall explained variance of 57.80% and showed all items to load highly on that factor (positive loadings for femininity items: $\geq$ 0.70; negative loadings for masculinity items: $\leq$ −0.67). Taken together, a one-factor solution was indicated. Factor, pattern, and structure matrix for the exploratory factor analysis and factor loadings for the confirmatory factor analysis can be found in Table B1 in Appendix B in Supplementary Material.

Group Differences

Table 3 shows overall scale means, average scores for each sexual-orientation group, and statistical tests. As expected, lesbians scored lower on TMF femininity and higher on TMF masculinity than bisexual or straight women. All differences between groups were statistically significant (based on a Scheffé test), except that bisexual women did not score significantly higher than straight women on masculinity. On the BSRI, no significant differences between groups were obtained. In contrast, regarding gender-role behavior and childhood behavior, expected differences between lesbians and straight women were obtained. Similarly, the implicit association of self with feminine was stronger in straight women than lesbians, confirming expectations.

Bivariate Correlations

Table 2 shows bivariate correlations, along with internal consistencies. Internal consistencies of all measures were excellent, with the lowest score obtained for BSRI masculinity. A noteworthy correlation was a strong negative one between the TMF factors masculinity and femininity, suggesting that a one-dimensional measure could be sufficient. In line with the large negative correlation, people who judged themselves as "moderately feminine" (i.e., ticked the value 4) tended to also judge themselves as "moderately masculine" (i.e., ticked 4 again). Hence, we recoded all masculine items and then averaged all items of the TMF to obtain a supplementary measure, TMF total. TMF masculinity and femininity correlated in the expected direction with all other measures except for BSRI masculinity. BSRI masculinity did not correlate significantly with any other measure, suggesting that it measured something different from all other measures of masculinity in the study. All other correlations were in the expected direction. Of particular interest, the implicit association of self-feminine correlated positively with TMF femininity and negatively with TMF masculinity, as expected. Similar, but somewhat weaker relations were obtained between the IAT and most other measures.

Predicting Sexual Orientation

In order to test whether lesbians, bisexual, and straight women would be classified correctly based on the different measures of masculinity-femininity, we carried out an ordinal regression analysis. As predictor variables, the masculinity and femininity scores of BSRI, GRB, and CGRB were entered. In addition, TMF total and the IAT effect were used as predictors. The overall model was statistically significant, $\chi_{(8)} = 72.01$, $p < 0.001$, Nagelkerke's $R^2 = 0.49$. The significant predictors were TMF total scores [$B = 1.17$, $SE = 0.27$, $\chi^2_{(1)} = 19.30$, $p < 0.001$] and masculine everyday behavior [$B = -0.69$, $SE = 0.27$, $\chi^2_{(1)} = 6.65$, $p = 0.01$]. None of the other predictors was significant, ps > 0.21. Thus, based on their self-assessment on the TMF as masculine-feminine and based on the masculine everyday behaviors participants said they carried out, they could be classified quite well as lesbians, bisexual, or straight women.

Mediation Analyses

Based on the regression approach suggested by Hayes (2013), we tested whether there are indirect effects of the BSRI and GRB dimensions on sexual orientation via the respective TMF dimensions. Because this approach needs a continuous dependent variable, in contrast to all other analyses in the present paper, we did not use the classification as lesbian, bisexual, or straight in this case, but the continuous Kinsey-like scale with scores ranging from 1 to 7. **Figures 2**, **3** summarize the findings. Statistically significant effects of BSRI femininity and GRB femininity on TMF femininity were observed, and also of GRB masculinity and of BSRI masculinity (by trend) on TMF masculinity. TMF masculinity and femininity were related with sexual orientation in expected ways (in line with the findings reported in **Table 3**). Bootstrapping analyses, using 10,000 Bootstrap re-samples, demonstrated that the indirect effects of BSRI femininity, GRB femininity, and BSRI femininity on sexual orientation via the TMF were statistically significant (i.e., none of the bias-corrected 95% confidence intervals included 0). The indirect effect of BSRI masculinity via TMF masculinity missed

TABLE 3 | Overall Scale Means (with *SD*) and Means per Group, with Statistical Test of Difference (all *df* = 2, 123; with effect size; Tukey HSD) and Correlation with the Sexual Orientation Scale in Study 1.

	M	Lesbians ($n = 47$)	Bisexual women ($n = 32$)	Straight women ($n = 47$)	Group differences (R_p^2)
TMF-F	4.24 (1.35)	3.20 (1.05)	4.47 (1.00)	5.13 (1.12)	$F = 40.19$, $p < 0.001$ (0.40) L < B < S
TMF-M	3.49 (1.22)	4.36 (0.90)	3.24 (1.02)	2.78 (1.09)	$F = 30.31$, $p < 0.001$ (0.33) L > B = S
BSRI-F	5.24 (0.75)	5.09 (0.71)	5.38 (0.85)	5.30 (0.69)	$F = 1.63$, $p = 0.19$, *ns*
BSRI-M	4.67 (0.73)	4.69 (0.75)	4.72 (0.83)	4.61 (0.65)	$F < 1$, *ns*
GRB-F	4.38 (0.85)	4.16 (0.80)	4.45 (0.83)	4.57 (0.89)	$F = 2.94$, $p = 0.06$ (0.05) L < S
GRB-M	4.22 (0.86)	4.73 (0.79)	4.09 (0.83)	3.78 (0.68)	$F = 19.29$, $p < 0.001$ (0.24) L > B = S
CGRB-F	3.97 (1.63)	2.93 (1.55)	4.28 (1.37)	4.79 (1.30)	$F = 21.40$, $p < 0.001$ (0.26) L < B = S
CGRB-M	4.29 (1.52)	5.24 (1.54)	3.98 (1.25)	3.56 (1.16)	$F = 19.93$, $p < 0.001$ (0.25) L > B = S
IAT effect	0.67 (0.31)	0.59 (0.31)	0.65 (0.31)	0.76 (0.30)	$F = 3.81$, $p < 0.03$ (0.06) L < S

All scales theoretically range from 1 to 7, except for the IAT effect that is similar to an effect size of the stronger self-feminine as opposed to self-masculine association. Abbreviations of tests: TMF, Traditional Masculinity-Femininity; BSRI, Bem Sex Role Inventory; GRB, Gender-Role Behavior; IAT effect: Differences in the implicit association test (IAT) between the self-masculine/others-feminine and the self-feminine/others-masculine task. Endings indicate masculinity (–M) und femininity (–F) scales. Higher scores indicate higher masculinity on masculinity scales and higher femininity on femininity scales. Abbreviations of groups: L, Lesbians; B, bisexual women; and S, straight women.

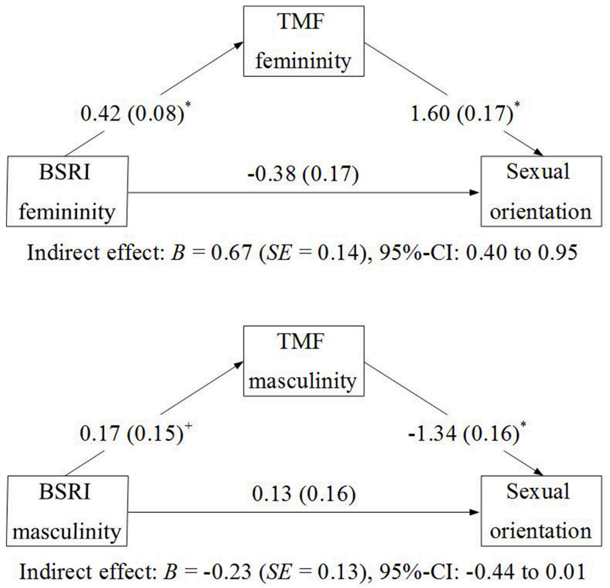


FIGURE 2 | Mediation of the relation between BSRI and sexual orientation by the TMF.

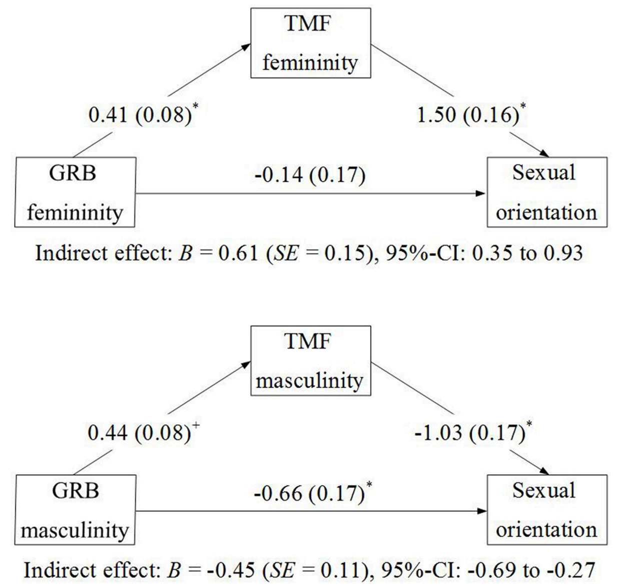


FIGURE 3 | Mediation of the relation between GRB and sexual orientation by the TMF.

the preset criterion of statistical significance. Only one direct effect was significant in addition to the indirect effect: Whereas all other findings were in line with the interpretation of full mediation via the TMF, masculine everyday behavior was still related to sexual orientation when the TMF was included in the equation. This suggests that the TMF mediated the relationship between sexual orientation and masculine behavior only partially.

Summary of Findings

In Study 1, we found that the reliabilities of both the femininity and the masculinity subscales of the TMF were high. Moreover, they correlated so strongly (in a negative way) that one may also conceive of the scale as one-dimensional, ranging from masculinity to femininity. We found several pieces of evidence for the validity of the scale. First, it correlated in the expected directions with all other measures of masculinity and femininity that we used, except for BSRI masculinity, which largely confirms Hypothesis 7. Feminine traits as well as masculine and feminine behaviors can be predicted quite well from scores on the TMF. The strongest correlations were obtained with self-rated childhood gender conformity. Notably, confirming Hypothesis 6, correlations with an implicit measure of one's self-feminine vs. self-masculine association were in the expected order of magnitude (e.g., Hofmann et al., 2005) and higher than those of the implicit measure with any of the trait or behavior self-ratings. Additionally, the TMF was related to participants' sexual orientation more strongly than any other measure (see ANOVA results in **Table 3**), with lesbians reporting lower femininity and higher masculinity than bisexual or straight women (confirming Hypothesis 3a and b). When predicting participants' sexual orientation from the masculinity and femininity measures, neither feminine, nor masculine traits, nor feminine everyday behavior, nor the self-feminine association contributed. Instead, confirming Hypothesis 8, masculine everyday behavior and the TMF were able to predict participants' sexual orientation very well, attesting to the usefulness of two rather new conceptualizations of measuring masculinity and femininity.

Mediation analyses were in line with the idea that feminine traits and feminine everyday behavior differ by sexual orientation because of a globally more feminine gender-role self-concept. This confirms Hypothesis 5. Masculine traits also tend to differ by sexual orientation because of lesbians' globally more masculine gender-role self-concept. Further, masculine everyday behavior also differs by sexual orientation because of lesbians' globally more masculine gender-role self-concept, but a direct effect of masculine behavior on sexual orientation remained. A speculative explanation for the latter finding is that it may depend partly on the gender of one's relationship partner which behaviors one carries out. For example, given that couples typically divide housework in ways mirroring traditional gender roles (e.g., Croft et al., 2014; Steffens and Viladot, 2015), a woman considering herself rather feminine may mow the lawn more often when she is in a relationship with a woman than with a man. In other words, in addition to personal preferences, the presence or absence of other-gender people in the household who choose to take care of certain chores may determine which chores one does (i.e., typically male everyday behaviors if no man is around).

STUDY 2

The aim of Study 2 was to replicate and extend Study 1's findings. We used data of a research project on social perception. As in Study 1, we used a known-groups approach, this time contrasting lesbians, gay men, and straight women and men.

With the exception of small adjustments, gender-related scales were identical to Study 1. However, this time we used a different adjective-based instrument than the BSRI, namely the GEPAQ, the German version (Runge et al., 1981) of the Extended Personal Attributes Questionnaire (Spence et al., 1978). For determining criterion validity, we also focused on other features. Participants were instructed to provide information regarding frequency of contact with lesbians/gay men and straight people. Moreover, characteristics of participants' voice pitch were collected as well as evaluations from independent judges on whether participants' voices sounded straight or gay/lesbian and whether their faces looked straight or gay/lesbian. In order to determine the TMF's test-retest reliability, we re-invited male participants after 1 year (for female participants no contact data were available).

We expected highest masculinity/lowest femininity scores for straight men, followed by gay men, lesbians, and straight women, implying lowest masculinity/highest femininity for straight women (Hypotheses 2, 3, and 4). We expected gender-related characteristics to correlate moderately with the TMF (Hypothesis 7) and we assumed the TMF to predict sexual orientation better than the other gender-related scales (Hypothesis 8). Furthermore, we assumed that participants with higher gender-conform scores on the TMF would report less contact with lesbians and gay men (Hypothesis 10), would show rather gender stereotypical voice pitch characteristics (Hypothesis 11), and would be more likely to be rated as straight (Hypothesis 9). A moderate 1-year reliability was expected (Hypothesis 12) as well as a moderate predictive validity for the second measurement of gender-related features (Hypothesis 13).

Method

Participants

Overall 111 German participants attended the study at the first measurement point. Their age ranged from 19 to 30 years ($M = 24.2$, $SD = 2.5$). Participants were recruited at the University of Jena, the Technical University of Berlin, and on lesbian/gay dating websites. Based on their Kinsey-like scale scores, 15 participants who rated themselves as bisexual were excluded from further analyses because of the small group size. Among the remaining 96 participants, there were 24 lesbians (Kinsey scores: 1–2), 21 straight women (6–7), 25 gay men (1–2), and 26 straight men (6–7). Most participants were well educated, 60% possessing a university entrance qualification and 35% a university degree. As a *post-hoc* power analysis indicated, given the sample size and $\alpha = 0.05$, between medium (0.25) and large (0.40) effects of $f = 0.35$ could be detected in the 2 × 2 ANOVAs below with a statistical power of $1 - \beta = 0.95$.

A total of 37 men attended the post-test. According to their Kinsey-like scale scores 18 identified as gay (1–2) and 19 as straight (6–7). Between those attending the post-test and those who did not, merely one difference was significant after adjusting the significance level for multiple tests. The retest-group reported less contact with straight men during the first data collection [$M_{retest} = 5.76$, $M_{no-retest} = 6.53$, $t_{(49.47)} = 3.33$, $p = 0.002$].

Materials

The same measures as in Study 1 were used in the following manner. Because the femininity and masculinity subscales of the TMF were highly correlated, as were subscales of the Childhood Gender-Role Behavior Scale, they were combined to form one dimension each [TMF: 1 (*very masculine*), to 7 (*very feminine*); CGRB: 1 (*I strongly disagree*), 5 (*I strongly agree*)]. Thus, the 6-item-version of the TMF was used. High values on CGRB indicated a high degree of gender conformity. Gender Role Behavior was assessed with a 6-point-scale this time and sexual orientation was measured with one item on a 7-point Kinsey-like scale [("Regarding sexual orientation, I identify as…"); 1 (*exclusively lesbian/gay*), 7 (*exclusively straight*)]. Moreover, we included the following measures.

German extended personal attributes questionnaire

We used the German version (Runge et al., 1981) of the Extended Personal Attributes Questionnaire (Spence et al., 1978). It consists of two independent scales measuring gender-related personality traits. The instrumentality scale (GEPAQ-M) contained eight items describing behaviors more socially desirable for men (e.g., independent), the expressiveness scale (GEPAQ-F) comprised eight items more socially desirable for women (e.g., emotional). Participants were instructed to indicate on a 5-point Likert-type scale (1 = non independent/not emotional, 5 = very independent/very emotional) the extent to which they felt each item described them.

Contact measures

In order to estimate the composition of participants' social environment, we measured current contact to same-gender lesbian/gay and straight people with one item each. The participants should "indicate how often you have contact to homosexual and heterosexual women/men" on a 7-point scale ranging from 1 (*never*) to 7 (*always*).

Voice pitch characteristics

To describe participants' voice pitch (i.e., the auditory correlate of fundamental frequency) distributions in spontaneous speech, we used three measures. Mean fundamental frequency (f0) indicates the average voice pitch, f0 standard deviation is a measure for voice pitch variability, and f0 range is used to evaluate voice pitch range. For computing f0 range, we computed the difference between the f0 97.5th percentile (estimator of the upper voice pitch boundary) and f0 2.5th percentile (estimator of the lower voice pitch boundary).

Perceived straightness

Participants' voices, facial photographs, and the combinations of both voices and faces had been rated as either "heterosexual" or "homosexual" by 101 judges (65 women, 31 men; age $M = 28.0$), participating in a different study (for details see Kachel et al., unpublished manuscript). To receive a relative measure of "heterosexual" judgments, all "heterosexual" responses were summed for each participant and divided by the number of judgments. Hence, higher scores indicate higher perceived straightness.

Procedure

At first measurement, participants filled out an online questionnaire in which all psychological and sociodemographic characteristics were collected. The order of psychological instruments was TMF, CGRB, contact to girls and boys during childhood, GRB, GEPAQ, Kinsey-like scale, and finally current contact to same-gender lesbians/gay men and straight people. In the second step, they were invited to a speech lab to provide recordings of spontaneous spoken speech and text reading as well as a photograph of their face. The sampling of women took place in a phonetic laboratory in the Center of General Linguistics in Berlin and was done by a female investigator, whereas the sampling of men took place at a phonetic laboratory of the University of Jena and was done by a male investigator. Voice pitch characteristics were measured on the basis of spontaneous speech. In the last step we asked 101 judges to rate speech recordings, facial photographs, and the combination of both dichotomously regarding sexual orientation for a randomly selected subset of 18 lesbians, gay men, straight women, and men, respectively (Kachel et al., unpublished manuscript). For the rating of speech recordings, we used the same read sentence for all target persons ("It has been quite a long day," German: "*Der Tag ist sehr lang geworden.*") in order to hold the conditions constant for every target and to control for the phonetic composition of the utterance.

Male participants were re-invited after 1 year to the phonetic laboratory of the University of Jena. Before speech recordings they were asked to fill out an online questionnaire containing several gender-related scales including the 6-items version of the TMF, the GEPAQ-M, and the GEPAQ-F.

Results

All results refer to the first measurement except for those that are explicitly indicated to belong to second measurement.

Factor Analysis

In order to test whether the TMF scale is one-dimensional, an explorative factor analysis with PAF was conducted. It replicated all findings of the pilot study. In detail, a KMO criterion of 0.86 indicated that the sample was appropriate. All items were suitable for factor analysis (item-specific KMO values >0.81; commonalities:0.54–0.83). According to a graphical scree-plot analysis, a one-factor solution was confirmed. There was a steep decline of explained variance from factor one (71%) to factor two (13%). Each item was represented very well by this factor (loadings >0.73).

An additional exploratory factor analysis with PAF of participants at second measurement replicated the findings indicating a one-dimensional factor structure. In detail, a KMO criterion of 0.76 indicated that the sample was appropriate. All items were suitable for factor analysis because of item-specific KMO values >0.69 and moderate to high commonalities (0.42–0.69). The one-factor solution was confirmed by graphical scree-plot analysis. There was a steep decline of explained variance from factor one (60%) to factor two (14%). Each item was represented very well by this factor (loadings >0.65).

Differences on Gender-Related Scales Based on Gender and Sexual Orientation

Which of the gender-related instruments are able to predict a person's gender and sexual orientation? In order to answer this question, for all gender-related instruments separate 2 × 2 ANOVAs with the two between-subject factors gender and sexual orientation were computed. Simple-effects tests with Bonferroni adjustment were added. **Table 4** shows main and interaction effects as well as mean scores for all gender-related instruments separately for lesbians, straight women, gay, and straight men.

On the TMF, we found an interaction of gender and sexual orientation, $F_{(1,\ 92)} = 21.42$, $p < 0.001$, $\eta_p^2 = 0.19$, as well as a main effect of gender $F_{(1,\ 92)} = 100.54$, $p < 0.001$, $\eta_p^2 = 0.52$. Both effects explained more variance in the TMF than in all other gender-related instruments in this study. Because straight women and men conform to gender roles more than lesbians/gay men, stronger gender differences should be expected between straight women and men than between lesbians and gay men. Hence, comparing lesbians and gay men constituted a stricter test of all scales. Although the TMF mean differences between straight women and men were more distinct ($\Delta M = 2.85$), lesbians and gay men significantly differed, too ($\Delta M = 1.05$). In short, the TMF showed the expected mean differences between all groups, it was the only scale in this study that was able to detect differences between lesbians and gay men, and it showed the largest mean difference between straight women and men.

Furthermore, the TMF differentiated the groups as expected (see **Figure 4**). Lesbians and straight women were on average clearly located on the scale's side that is associated with femininity (scores > 4) and gay and straight men's mean values were connected to masculinity (scores < 4). Additionally, the TMF was best in predicting gender on the basis of scale scores as can be seen in **Table 5** in which results of binary logistic regression models for all gender-related scales are shown. Correct gender classification rate for the TMF was 80%. Almost identical percentages of women and men were correctly classified. Compared to all other measures under investigation, the TMF seemed to be the most precise instrument to differentiate between women and men regardless of their sexual orientation.

Replication of Findings from the Female Sample

Group differences in women's sample

Regarding TMF, group differences in women's sample were already mentioned above. As in Study 1, straight women described themselves as more feminine compared to lesbians on the GRB-F. However, in contrast to Study 1, other gender-related scales (GRB-M and CGRB) were not able to differentiate women regarding their sexual orientation (see **Table 4**). Means were particularly close together for adjective-based gender-related instruments such as the GEPAQ.

Bivariate correlations

Reliabilities and correlations on all gender-related instruments can be seen in **Table 6**. Three out of five correlations with the TMF were significant. Besides the GRB-F there was also a correlation with gender-conforming childhood-experiences

TABLE 4 | Group-Specific Means (with *SD*) on Gender-Related Scales and ANOVA Results regarding Sexual Orientation and Gender in Study 2 at First Measurement.

	Women		Men		SexOr (F, p, η^2)	Gender (F, p, η^2)	Gender × SexOr (F, p, η^2)
	L[1]	S[2]	G[3]	S[4]			
TMF	^{a}4.54 (1.15)	^{b}5.36 (0.72)	^{c}3.49 (0.87)	^{d}2.51 (0.98)	0.15, 0.703, 0.00	100.54, < 0.001, 0.52	21.42, < 0.001, 0.19
GEPAQ-M	^{a}3.29 (0.45)	^{a}3.23 (0.40)	^{a}3.31 (0.70)	^{a}3.41 (0.55)	0.02, 0.886, 0.00	0.79, 0.376, 0.01	0.47, 0.494, 0.01
GEPAQ-F	^{a}4.04 (0.55)	^{a}4.18 (0.51)	^{a}4.06 (0.48)	^{b}3.68 (0.48)	1.48, 0.227, 0.02	5.36, 0.023, 0.06	6.35, 0.013, 0.07
GRB-M	^{a}3.59 (0.78)	^{a}3.45 (0.79)	^{a}3.42 (0.66)	^{b}4.17 (0.78)	3.08, 0.082, 0.03	1.94, 0.167, 0.02	6.41, 0.013, 0.07
GRB-F	^{a}3.91 (0.76)	^{b}4.63 (0.45)	^{a}3.57 (0.77)	^{c}3.17 (0.78)	1.03, 0.313, 0.01	37.11, 0.001, 0.29	14.42, < 0.001, 0.14
CGRB	^{a}3.04 (1.10)	^{a}3.38 (0.92)	^{a}3.22 (0.67)	^{b}4.32 (0.52)	18.05, < 0.001, 0.16	11.00, 0.001, 0.11	4.99, 0.028, 0.05

TMF: 1-7, Traditional Masculinity-Femininity; GEPAQ: 1-5, German Extended Personality Attributes Questionnaire; GRB: 1-6, Gender-Role Behavior; and CGRB: 1-5, Childhood Gender-Role Behavior. Endings indicate masculinity (–M) und femininity (–F) scales. Higher scores indicate higher masculinity on masculinity scales, higher femininity on femininity scales and TMF, and higher gender-conformity on CGRB. According to a Levene test, the assumption of homogeneity of variance was violated for GRB-F and CGRB. Superscripted letters in mean columns refer to groups based on simple-effect findings. Groups sharing the same letter do not differ significantly from each other at $\alpha \leq 0.05$.

[1] *Lesbians n = 25.*

[2] *Straight women n = 26.*

[3] *Gay men n = 24.*

[4] *Straight men n = 21.*

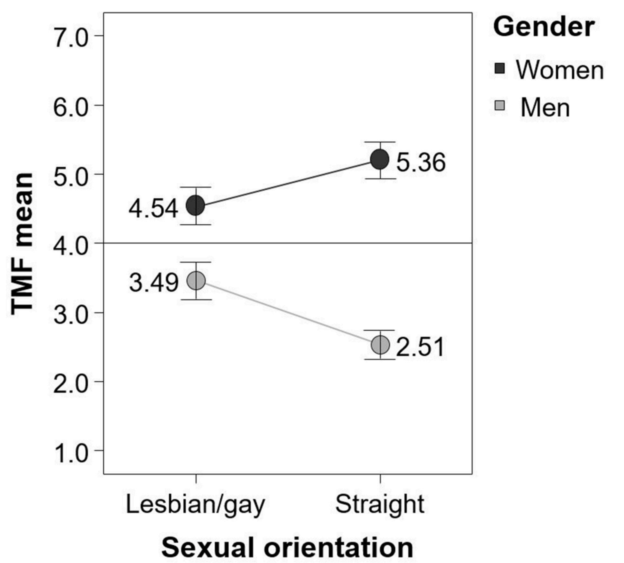


FIGURE 4 | Mean TMF scores separately for gender and sexual orientation. Error bars represent standard errors of means.

(CGRB) and with the exchanged adjective-based masculinity-scale (GEPAQ-M). The correlations for the first two instruments were in the expected direction: The more feminine the women rated themselves on the TMF, the higher their scores on behavior-based femininity (GRB-F) and childhood gender-conformity (CGRB). However, the TMF correlated positively with the GEPAQ-M, which is counterintuitive. We believe that this attests to deficiencies in the GEPAQ-M, along with its low reliability. Moreover, after adjusting the significance level according to the Bonferroni formula, the correlation was not significant anymore.

Predicting sexual orientation

Can the TMF predict women's sexual orientation better than other measures? We added the TMF in the last step of a binary regression model. Results can be seen in **Table 7**. In contrast to Study 1, the TMF did not outperform all other measures. Only the GRB-F was found to predict women's sexual orientation. However, when GRB-F was not included in the regression model, the TMF was the only significant predictor of sexual orientation in the model, $B = 1.25$, $SE = 0.50$, $\chi^2_{(1)} = 6.19$, $p = 0.013$.

TABLE 5 | Results of Binary Logistic Regression Models in Predicting Participants' Gender based on Different Gender-Related Instruments in Study 2 at First Measurement.

	Percentage of correct classifications			B	SE	$\chi^2_{(1)}$	p
	Overall	Men	Women				
TMF	80.2	80.4	80.0	1.67	0.33	25.50	<0.001
GEPAQ-M	53.1	82.4	20.0	−0.35	0.39	0.79	0.374
GEPAQ-F	60.4	68.6	51.1	0.90	0.42	4.64	0.031
GRB-M	49.0	64.7	31.1	−0.36	0.27	1.81	0.179
GRB-F	70.8	68.6	73.3	1.42	0.32	19.71	<0.001
CGRB	60.4	72.5	46.7	−0.69	0.24	8.23	0.004

TMF, Traditional Masculinity-Femininity; GEPAQ, German Extended Personality Attributes Questionnaire; GRB, Gender-Role Behavior; and CGRB, Childhood Gender-Role Behavior. Endings indicate masculinity (–M) und femininity (–F) scales.

Comparisons within Men

The same analyses were computed for the male subsample.

Group differences

As indicated in **Table 4**, all differences in the male subsample were in the expected directions. Straight men showed higher masculinity/lower femininity on each gender-related instrument than gay men except for the GEPAQ-M, where no significant difference was detected. The TMF ($\Delta M = 1.05$) and the CGRB ($\Delta M = 1.10$) were similarly able to predict sexual orientation.

TABLE 6 | Reliabilities and bivariate correlations of gender-related scales for women and men in Study 2 at first measurement.

	TMF	GEPAQ-F	GEPAQ-M	GRB-F	GRB-M	CGRB
TMF	0.86/0.89	(0.28)	0.34	0.47	(−0.27)	0.54
GEPAQ-F	0.59	0.77/0.64	(0.04)	(0.29)	(0.21)	(0.27)
GEPAQ-M	−0.38	(−0.21)	0.51/0.73	(−0.03)	(−0.09)	(0.12)
GRB-F	0.31	0.46	(0.19)	0.88/0.91	(0.29)	0.47
GRB-M	−0.45	−0.26	0.42	(0.20)	0.83/0.88	(−0.14)
CGRB	−0.55	−0.41	(0.18)	−0.35	0.37	0.82/0.73

Correlations for women sample are presented above, for men sample below the diagonal. Internal consistencies are depicted in the diagonal with the values before the slash referring to women and after the slash referring to men. All correlations are statistically significant at $\alpha \leq 0.05$ except for those in parentheses. Abbreviations: TMF, Traditional Masculinity-Femininity; GEPAQ, German Extended Personality Attributes Questionnaire; GRB, Gender-Role Behavior; and CGRB, Childhood Gender-Role Behavior. Endings indicate masculinity (–M) und femininity (–F) scales. Higher scores indicate higher masculinity on masculinity scales, higher femininity on femininity scales and TMF, and higher gender-conformity on CGRB.

At second measurement, gay and straight men differed more strongly on the TMF [$M_{gay} = 3.85$, $M_{straight} = 2.60$, $t_{(35)} = 4.70$, $p < 0.001$]. However, in contrast to the first measurement the GEPAQ-F was not able to discriminate between both groups, $M_{gay} = 4.02$, $M_{straight} = 3.68$, $t_{(35)} = 1.83$, $p = 0.075$. The GEPAQ-M remained non-significant, $M_{gay} = 3.46$, $M_{straight} = 3.56$, $t_{(35)} = -0.51$, $p = 0.61$.

Bivariate correlations

All correlations with the TMF were significant (all $|r| > 0.31$, all $p < 0.028$) and in the expected directions (see **Table 6**).

Predicting sexual orientation

As for the female subsample, the TMF did not predict sexual orientation better than other measures when it was added in the last step of a binary regression model (see **Table 8**). CGRB and GRB-M were the measures most closely related to sexual orientation. This could be interpreted as suggesting that TMF does not contribute at all to explaining sexual orientation. Moreover, one could be interested in the direct comparison of TMF and GEPAQ in explaining sexual orientation. To answer these questions, in a supplementary binary regression model, only adjective-based scales were included as predictors. In that analysis, TMF was the only significant predictor of sexual orientation, $B = -0.89$, $SE = 0.41$, $\chi^2_{(1)} = 4.61$, $p = 0.032$. Taken together, CGRB and GRB-M predicted sexual orientation best, and TMF predicted sexual orientation better than GEPAQ.

Relations with Criterion Characteristics

We collected data on several psychological and acoustic criterion characteristics. We computed bivariate correlation coefficients for the TMF with these characteristics in order to test the criterion validity of TMF separately for women (see **Table 9**) and men (see **Table 10**). Additionally, correlations for all other gender-related scales included in Study 2 were computed as a comparison.

The more gender-conform women and men rated themselves on the TMF, the more likely they were perceived as straight

TABLE 7 | Stepwise, logistic regression analysis for predicting women's sexual orientation based on gender-related scales in study 2 at first measurement.

	Step 1	Step 2	Step 3
Predictors:	***B (SE)*** **Exp(B)**	***B (SE)*** **Exp(B)**	***B (SE)*** **Exp(B)**
GEPAQ-F	[0.50 (0.58) 1.66]	[0.39 (0.76) 1.48]	[0.72 (0.92) 2.05]
GEPAQ-M	[−0.37 (0.72) 0.69]	[−0.86 (0.87) 0.43]	[−0.95 (0.94) 0.39]
GRB-F		3.01 (0.97) 20.30	3.69 (1.38) 40.15
GRB-M		−1.13 (0.56) −0.52	[−1.43 (0.80) 0.24]
TMF			[0.35 (0.66) 1.42]
CGRB			[−0.90 (0.58) 0.41]
	$\chi^2_{(2)} = 1.01$, $p = 0.605$, $R^2 = 0.03$, 56%	$\chi^2_{(2)} = 18.68$, $p < 0.001$, $R^2 = 0.47$, 73%	$\chi^2_{(2)} = 2.99$, $p = 0.224$, $R^2 = 0.53$, 78%

Chosen method was "Forward: Wald" in each block. R^2 means Nagelkerke's R^2. Percentage values refer to correctly classified lesbian and straight women. All correlations are statistically significant at $\alpha = 0.05$ except for those in brackets. TMF, Traditional Masculinity-Femininity; GEPAQ, German Extended Personality Attributes Questionnaire; GRB, Gender-Role Behavior; and CGRB, Childhood Gender-Role Behavior. Endings indicate masculinity (–M) und femininity (–F) scales. Higher scores indicate higher masculinity on masculinity scales, higher femininity on femininity scales and TMF, and higher gender-conformity on CGRB.

TABLE 8 | Stepwise, logistic regression analysis for predicting men's sexual orientation based on gender-related scales in study 2 at first measurement.

	Step 1	Step 2	Step 3
Predictors:	***B (SE)*** **Exp(B)**	***B (SE)*** **Exp(B)**	***B (SE)*** **Exp(B)**
GEPAQ-F	−1.82 (0.72) 0.16	[−0.71 (0.88) 0.49]	[0.64 (1.38) 1.90]
GRB-F		−1.25 (0.61) 0.29	[−1.09 (0.89) 0.34]
GRB-M		1.64 (0.55) 5.17	1.90 (0.86) 6.69
TMF			[−0.25 (0.66) 0.78]
CGRB			3.77 (1.28) 43.39
	$\chi^2_{(1)} = 8.17$, $p = 0.004$, $R^2 = 0.20$, 64%	$\chi^2_{(2)} = 12.57$, $p = 0.002$, $R^2 = 0.47$, 71%	$\chi^2_{(2)} = 19.50$, $p < 0.001$, $R^2 = 0.73$, 86%

Chosen method was "Forward: Wald" in each block. R^2 means Nagelkerke's R^2. Percentage values refer to correctly classified gay and straight men. All correlations are statistically significant at $\alpha \leq 0.05$ except for those in brackets. TMF, Traditional Masculinity-Femininity; GEPAQ, German Extended Personality Attributes Questionnaire; GRB, Gender-Role Behavior; and CGRB, Childhood Gender-Role Behavior. Endings indicate masculinity (–M) und femininity (–F) scales. Higher scores indicate higher masculinity on masculinity scales, higher femininity on femininity scales and TMF, and higher gender-conformity on CGRB.

based on voices, faces, and the combination of both ($|r| > 0.31$) however, the correlation for perceived straightness based on voice for women was only by trend). In contrast to men, all voice pitch characteristics correlated significantly with the TMF for women ($r > 0.32$). All correlations were in the expected direction: The higher women spoke on average and the higher their voice pitch range and variability, the more likely they rated themselves as feminine. In contrast, one contact measure showed a significant correlation for men but not for women: The less contact men reported to gay men, the more masculine they rated themselves on the TMF ($r = -0.35$).

TABLE 9 | Bivariate correlations of gender-related instruments and criterion characteristics for women in study 2.

	TMF	GEPAQ		GRB		CGRB
		F	M	F	M	
Contact to lesbians/ gay men	(−0.27)	(−0.03)	(0.20)	(−0.12)	0.30	(−0.10)
Contact to straight wo/men	(0.20)	0.32	(−0.08)	(0.24)	(−0.01)	(0.22)
Voice pitch average	0.41	(0.24)	(0.07)	0.37	(−0.14)	0.46
Voice pitch variability	0.32	(0.14)	(−0.10)	(0.26)	(−0.08)	0.35
Voice pitch range	0.34	(0.19)	(−0.04)	(0.29)	(−0.06)	0.39
Perceived straightness in voice	(0.31)	(0.05)	(−0.01)	(0.20)	(−0.16)	(0.30)
Perceived straightness in face	0.57	(0.09)	(0.03)	(0.29)	(−0.20)	(0.23)
Perceived straightness in voice + face	0.55	(0.11)	(0.01)	(0.31)	(−0.17)	(0.24)

All correlations are statistically significant at $\alpha \leq 0.05$ except for those in parentheses. Abbreviations for column headings: TMF, Traditional Masculinity-Femininity; GEPAQ, German Extended Personality Attributes Questionnaire; GRB, Gender-Role Behavior; and CGRB, Childhood Gender-Role Behavior. Endings indicate masculinity (–M) und femininity (–F) scales. For gender-related instruments higher scores indicate higher masculinity on masculinity scales, higher femininity on femininity scales and TMF, and higher gender-conformity on CGRB. For criterion characteristics higher scores indicate more frequent contact to lesbians/gay men and straight wo/men, higher voice pitch characteristics, and higher perceived straightness.

The TMF showed 9 out of 16 possible significant correlations which is more than any other gender-related scale. CGRB followed with 6 out of 16 possible significant correlations. Hence, the TMF showed higher convergent validity than the other gender-related scales.

Test-Retest Reliability and Predictive Validity

Table 11 contains findings regarding test-retest reliability and predictive validity. According to the intercorrelation of TMF scores at first and second measurement, 1-year reliability for the TMF was 0.75 and higher than for the GEPAQ-F, though inter-correlations for the GEPAQ-M were even higher than for the TMF. Hypothesis 12 was confirmed.

In order to test its predictive value, the TMF at the first measurement was correlated with GEPAQ-M and GEPAQ-F at the second measurement. As can be seen in **Table 11**, both correlations were significant, of moderate size, and in the expected directions, confirming Hypothesis 13.

Summary of Findings

In Study 2, we found that all TMF items loaded strongly on one single factor at first and second measurement, replicating the pilot study and confirming Hypothesis 1 again. The TMF showed sufficient reliabilities for women and men. Confirming Hypotheses 2, 3, and 4, the TMF turned out to be the best gender-related instrument for differentiating straight and gay men at first and second measurement and lesbians and straight women compared to all other scales used in Study 2 (see **Table 4**). In line with gender self-stereotyping and contradicting implicit gender inversion theory, gay men showed lower femininity/higher

TABLE 10 | Bivariate correlations of gender-related instruments and criterion characteristics for men in study 2.

	TMF	GEPAQ		GRB		CGRB
		F	M	F	M	
Contact to lesbians/ gay men	0.35	0.29	(−0.03)	(0.12)	(−0.08)	(−0.16)
Contact to straight wo/men	(−0.01)	(−0.20)	(00.07)	−0.42	(00.05)	(00.08)
Voice pitch average	(−0.06)	(−0.09)	(0.20)	(0.10)	(−0.01)	(0.00)
Voice pitch variability	(−0.27)	(−0.25)	(0.19)	(−0.17)	(0.23)	0.28
Voice pitch range	(−0.08)	(−0.22)	(0.12)	(−0.07)	(−0.10)	(0.05)
Perceived straightness in voice	−0.34	(−0.19)	(0.04)	(−0.04)	(0.14)	(0.22)
Perceived straightness in face	−0.38	−0.45	(0.17)	−0.47	(0.14)	0.39
Perceived straightness in voice + face	−0.47	−0.49	(0.17)	−0.42	(0.21)	0.48

All correlations are statistically significant at $\alpha \leq 0.05$ except for those in parentheses. Abbreviations for column headings: TMF, Traditional Masculinity-Femininity; GEPAQ, German Extended Personality Attributes Questionnaire; GRB, Gender-Role Behavior; and CGRB, Childhood Gender-Role Behavior. Endings indicate masculinity (–M) und femininity (–F) scales. For gender-related instruments higher scores indicate higher masculinity on masculinity scales, higher femininity on femininity scales and TMF, and higher gender-conformity on CGRB. For criterion characteristics higher scores indicate more frequent contact to lesbians/gay men and straight wo/men, higher voice pitch characteristics, and higher perceived straightness.

TABLE 11 | Reliabilities and correlations for gender-related measures between first (columns) and second (rows) Measurement in Study 3.

	Alpha	TMF[a]	GEPAQ-M[a]	GEPAQ-F[a]
TMF[b]	0.87	0.75	(−0.08)	0.49
GEPAQ-M[b]	0.73	−0.32	0.89	(−0.25)
GEPAQ-F[b]	0.75	0.35	(0.03)	0.65

Internal consistencies for second measurement are presented in the first column. Test-retest reliabilities are presented on the diagonal. All correlations are statistically significant at $\alpha \leq 0.05$ except for those in parentheses. Abbreviations: TMF, Traditional Masculinity-Femininity; GEPAQ, German Extended Personality Attributes Questionnaire. Endings indicate masculinity (–M) und femininity (–F) scales. Higher scores indicate higher masculinity on masculinity scales and higher femininity on femininity scales and TMF.

[a] *First measurement.*

[b] *Second measurement.*

masculinity than lesbians. The evidence for high incremental validity in predicting women's sexual orientation from Study 1 could not be replicated nor extended to men.

Whereas, lesbians and straight women differed descriptively, but not significantly in GRB-M (see **Table 4**), in the logistic regression analysis (see **Table 7**), GRB-M predicted women's sexual orientation in a significant way in Step 2, along with GRB-F. We assume that the inclusion of GRB-F in the regression model reduced apparent error variance and thus changed the relation between GRB-M and sexual orientation from descriptive to statistically significant. However, as GRB-M was again non-significant in Step 3 of the regression model, we suggest that masculine everyday behavior was not strongly related to sexual orientation in our women's sample. However, when including adjective-based instruments only, TMF predicted sexual orientation in women and men better than established adjective -based instruments.

Partially confirming Hypothesis 7, the TMF showed moderate correlations with some other gender-related scales. Importantly, the TMF was connected to multiple criterion characteristics for women (e.g., higher femininity was accompanied by more gender-conform voice pitch characteristics) and men (e.g., higher masculinity was associated with less frequent contact to gay men) and outperformed other gender-related scales.

The TMF revealed moderate test-retest-reliability and predictive validity confirming Hypotheses 12 and 13. Scores on the first TMF measurement predicted scores on GEPAQ-M and GEPAQ-F at second measurement.

GENERAL DISCUSSION

Gender research has developed many instruments to measure different aspects of self-ascriptions of gender stereotypical features, including attributes, behaviors, interests, and attitudes (Beere, 1990). Supplementing these scales, the TMF scale is designed as an instrument for globally assessing people's overall, or "core," masculinity-femininity. The TMF was shown to reliably measure an underlying, one-dimensional construct, and it was found to be a valid instrument for assessing masculinity-femininity because it (a) successfully differentiated between groups that were expected to differ (women vs. men, lesbians/gay men vs. straight women and men) and (b) it correlated moderately with other gender-related instruments, such as the Bem Sex Role Inventory (BSRI; Bem, 1974) and the German Extended Personal Attributes Questionnaire (GEPAQ; Runge et al., 1981). Whereas, some well-established, adjective-based scales (e.g., BSRI, GEPAQ) have shown shortcomings in differentiating women and men in recent years (Sczesny et al., 2004; Evers and Sieverding, 2014), our findings of consistent group differences support the TMF as a new tool for measuring gender-role self-concept.

Dimensionality of the TMF

In line with Choi and Fuqua (2003), high correlations between the separate TMF femininity and masculinity scales as shown in Study 1 suggest a bipolar, one-dimensional use of this instrument reflecting laypersons' ideas of masculinity and femininity as two extremes of one continuum. This is also in line with findings reported by Spence and Bruckner (2000, see also Sánchez and Vilain, 2012). All items were shown to load on one factor and represent a one-dimensional construct (masculinity-femininity). This finding should be not taken as hint that one-dimensional masculinity/femininity models generally outperform two dimensional ones (e.g., agency, communion; competence, warmth; instrumentality, expressivity), but that all TMF items appear to refer to the same underlying construct. Moreover, in spite of its brevity, the TMF showed high internal consistencies across all studies as well as satisfactory test-retest reliability (in a sample of men). However, the one-dimensionality of the TMF was demonstrated with participants identifying themselves as women or men. Possibly, the two-dimensional TMF version is superior than the one-dimensional version for samples that comprise a larger number of participants transgressing or rejecting the binary gender system (e.g., transgender and queer people). Future research is needed to clarify that question.

One could object against using the bipolar TMF scale that its midpoint is ambiguous. In other words: what does a score of "4" mean? One could imagine that people scoring either high or low on both dimensions would erroneously be treated as one group. However, according to the high correlations between the separate TMF masculinity and femininity scales (Study 1) and a supplementary graphical scatterplot analysis we did, we found no groups of high/high (i.e., androgyny) or low/low scorers (i.e., undifferentiated). Hence, it can be deduced that people in our samples who scored close to "4" believed themselves to be moderately feminine and masculine.

Contextualizing Validity Findings

In terms of validity, using a known-groups approach as an established psychological method for validity tests (e.g., Howitt and Cramer, 2008), the TMF repeatedly showed expected gender differences, with men scoring higher on masculinity and lower on femininity than women. With reference to sexual orientation, straight and bisexual women rated themselves higher on femininity and lower on masculinity than lesbians did (Study 1). Moreover, the TMF was the only gender-related scale used in the present study that distinguished straight men, gay men, lesbians, and straight women (from high masculinity/low femininity to low masculinity/high femininity, Study 2) which supports gender self-stereotyping rather than implicit gender inversion theory (Kite and Deaux, 1987). According to implicit gender inversion theory, gay men should have scored higher than lesbians on femininity and lower on masculinity, which was not the case in our sample. It appears that gay men and lesbians rather self-stereotype as men and women, respectively, and thus construct their self-concept in line with their gender group. Based on these findings, we conclude that the TMF's ability for determining gender and sexual orientation was generally high, and higher than that of all other gender-related measures investigated in the present studies. Finally, we found evidence for the idea that differences in "core" masculinity and femininity measured by the TMF underlie differences in lesbians' and gay men's vs. straight women and men's self-ascriptions in gender typicality measured by other scales, such as the BSRI (see Study 1). Hence, the TMF was shown to be a valid scale for assessing gender-role self-concept.

It was expected that the TMF would correlate moderately with other gender-related scales. That was the case for all gender-related scales in Study 1 where only a female sample was tested. This indicates that the TMF measures other aspects of people's conceptualizations of their own masculinity/femininity than the BSRI or the Gender-Role Behavior Scale (Athenstaedt, 2003) and complements them well. An explanation for this findings is that the TMF does not measure attributes associated with masculinity/femininity, but rather, these constructs themselves. Only correlations with the Childhood Gender-Behavior Scale were high, which could be due to selective memory recall and hence reflect current gender-related self-assessment (see Bailey and Zucker, 1995) measured with the TMF. Alternatively, the high correlation is due to actual gender differences during childhood, which would be a hint for constancy of conceptualizations of people's own masculinity/femininity. Correlations between the TMF and gender-related scales were smaller for a second sample of women (Study 2) which could be due to differences in sampling and substitutions of scales (e.g., instead of the Bem Sex Role Inventory, the Personal Attributes Questionnaire was used). Connected to that, the incremental validity of the TMF for predicting women's sexual orientation was demonstrated in Study 1 only. However, the male sample in Study 2 showed overall moderate correlations of the TMF and gender-related scales, but no additional ability of the TMF to predict sexual orientation. The fact that the TMF did not always demonstrate additional predictive value for explaining differences between groups does not indicate that it is superfluous. Rather, other facets of self-ascribed masculinity/femininity, such as everyday behavior, turned out to be highly capable of predicting sexual orientation as well. And the TMF predicted sexual orientation still better than established adjective-based instruments in women and men in Study 2 (which was demonstrated after excluding the most predictive scales).

To deal with a common critique that self-report instruments measure differences in social desirability rather than true differences, we used an implicit measure of women's self-feminine vs. self-masculine associations. Study 1 showed that the correlations of these associations were higher for the TMF than for self-ratings of traits or behaviors. This is a strong hint that the TMF is able to reflect "true" differences in core masculinity/femininity rather than social desirability only. It is also a substantive finding of the present studies that goes beyond mere scale validation.

In a similar vein, in order to test the criterion validity of the TMF, we selected several criterion characteristics which can be categorized into three groups (Study 2): These included contact to same-gender straight women/men and lesbians/gay men, voice pitch features, and assessment of sexual orientation by laypersons based on visual and auditory stimuli. Correlation analyses showed that gender-conformity on the TMF was significantly linked to perceived straightness for almost each presentation mode (voice, face, and the combination of both) for men and women. Moreover, higher femininity in women was associated with higher voice pitch features (average, variability, and range) and higher masculinity in men was connected to less contact to gay men. Compared to other gender-related scales, the TMF was superior in convergent validity. Taken together, self-ratings of masculinity/femininity go along not only with gender and sexual orientation differences, but also with differences in social behavior (i.e., contact to same-gender people differing in sexual orientation), with objective voice characteristics, and with assessments of sexual orientation based on facial and voice features. In sum, this indicates that the TMF measures something fundamental regarding gender-related self-assessment. It is also another substantive finding of the present studies that goes beyond mere scale validation. A limitation is that patterns of findings partially differed between women and men, and which specific criteria mattered in which sub-sample appeared a bit arbitrary (e.g., voice pitch features for women and contact variables for men). It appears that women and men express their masculinity/femininity in different ways, which is an interesting topic for future research.

Theoretical Considerations Regarding the TMF

One might assume that a one-item-measure could be sufficient for assessing masculinity/femininity by simply asking how masculine/feminine people believe themselves to be. We checked this idea in an exemplary fashion for Study 2 using the "I consider myself as…"-item for a comparative analysis because of highest corrected item-total correlations for the whole sample in the Pilot Study. However, in every case (determining and predicting gender and sexual orientation, convergent, and criterion validity), as a rule the TMF was better than the one-item-measure (e.g., compared to the one-item measure the TMF showed higher correlations for almost all gender-related measures in the male subsample except for GEPAQ-M where a higher correlation was found for the one-item measure). This is in line with state-of-the-art conceptions in psychological assessment that consider every item in a scale to be a piece of puzzle and hence uncover a different detail of a somewhat bigger picture (Bühner, 2010). Moreover, it is also consistent with Constantinople's (1973) view that the masculinity/femininity-construct is captured best when gender role adoption, preference, and identity are measured in conjunction.

The TMF is designed as a self-assessment instrument for masculinity-femininity on a rather global level with regard to two different respects. First, the TMF is based on a trait rather than a normative approach (see Thompson and Bennet, 2015) and conceptualizes masculinity-femininity as a long-term characteristic varying between people. However, it does not exclude variation on masculinity-femininity within a person depending on different social, temporal, or regional contexts. Its focus is on a trait-like (global) average score across contexts. Second, it is more global because it focuses on a higher-order masculinity-femininity construct which is beyond specific components such as traits, interests, physical characteristics, or attitudes, and asks for an aggregated self-assessment across these domains. The high test-retest reliability obtained over a 1-year period indicated stability rather than variance. However, it would be interesting to know which components mainly account for an individual's judgment of their own gender-related identity. The TMF could be a valuable instrument for future research dealing with that question.

In spite of this trait-like approach, the TMF is based on the idea that masculinity/femininity is socially determined (see Smiler, 2004). The scale is about how people relate or conform to social standards (how masculine/feminine do they believe themselves to be?), but not how they consider social norms to be appropriate for men and women (i.e., what people consider as masculine/feminine). To trigger a reference to social norms in the participants' minds when testing gender-role identity aspects, we used the term "traditionally" in the beginning of the corresponding items. However, the TMF does not measure if participants' conceptions of gender-role identity aspects correspond to traditional views. Thus, we concede that there could be variations in people's understanding of "traditionally" which could affect their self-evaluations. However, large differences are not likely because people within one culture know about traditional gender roles.

Because of the TMF's broader scope compared to established scales, such as the BSRI and PAQ, it is reasonable to be positive about the TMF's ability of measuring masculinity/femininity also in the future. Hence, it seems plausible that the problem of item aging is mitigated for the near future because of the more global wordings. Additionally, we are positive that the TMF can be used in different countries and cultures because of its global level of measurement. To date, the TMF has only been applied to one other German sample by Roth and Mazziotta (2015). They found that the TMF was moderately connected to different aspects of social identification with one's own gender in the expected directions for men and women. According to Leach et al. (2008), social identification is a multidimensional multicomponent higher order construct. The TMF was shown to be linked to almost all of its different components (individual self-stereotyping, in-group homogeneity, satisfaction, solidarity, and centrality) for women and men except for in-group homogeneity for men. Future research should provide evidence for the applicability in non-German samples.

Concluding Remarks

In a nutshell, as long as societies assume differences in interests, attitudes, clothing style, and behavior between women and men, we suggest that the TMF provides a valuable addition to researchers' toolbox. For example, are self-ratings on the TMF related to biological markers of masculinity-femininity such as waist-to-hip ratio and finger length (i.e., two-digit-four-digit ratio)? Do self-ratings on the TMF predict behaviors in which large gender differences have been observed, such as socio-sexuality or animal cruelty? Are self-ratings on the TMF related to performance in domains where gender differences are reliable, such as mental rotation? Finally, are self-ratings on the TMF related to personality traits in which gender differences have been observed, such as self-esteem and social dominance orientation? Generally, we believe that many different research questions related to gender-related self-assessments could benefit from using the TMF.

AUTHOR CONTRIBUTIONS

Substantial contributions to the conception or design of the work and the acquisition and analysis of the data: SK, MS; interpretation of data for the work: SK, MS, CN. Drafting the work or revising it critically for important intellectual content: SK, MS, CN. Final approval of the version to be published: SK, MS, CN. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved: SK, MS. We thank Kornelia Schertzl, Karoline Nestler, Dirk Hertrampf, Felicia Schuld, and Alexander Makosch for help with data collection, Susanne Fuchs, Stefanie Jannedy, and Joerg Dreyer for providing laboratories in the Zentrum fuer Allgemeine Sprachwissenschaft, Berlin, and Anders Sonderlund for language editing. Additionally, we thank Julia Scholz and the reviewers for critical and valuable comments on earlier versions of the manuscript. The TMF was originally developed by MS and Kornelia Schertzl.

FUNDING

The current research was partially funded by grants from the German Research Foundation (DFG, STE 938/10-2, FOR 1097, and STE 938/11-1).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00956

REFERENCES

Abele, A. E. (2000). "A dual impact model of gender and career related processes," in *The Developmental Social Psychology of Gender*, eds T. Eckes and H.-M. Trautner (Hillsdale, NJ: Erlbaum), 361–388.

Abele, A. E., and Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *J. Pers. Soc. Psychol.* 93, 751–763. doi: 10.1037/0022-3514.93.5.751

Alvesson, M. (1998). Gender relations and identity at work: a case study gender relations and identity at work: a case study of masculinities and femininities in an advertising agency. *Hum. Relat.* 51, 969–1005. doi: 10.1177/001872679805100801

Athenstaedt, U. (2000). Normative geschlechtsrollenorientierung: entwicklung und validierung eines fragebogens [Normative Gender Role Attitudes: The Development and Validation of a New Questionnaire]. *Zeitschrift Differentielle Diagnostische Psychologie* 21, 91–104. doi: 10.1024//0170-1789.21.1.91

Athenstaedt, U. (2003). On the content and structure of the gender role self-concept: including gender-stereotypical behaviors in addition to traits. *Psychol. Women Q.* 27, 309–318. doi: 10.1111/1471-6402.00111

Baeck, H., Corthals, P., and van Borsel, J. (2011). Pitch characteristics of homosexual males. *J. Voice* 25, 211–214. doi: 10.1016/j.jvoice.2010.10.019

Bailey, J. M., and Zucker, K. J. (1995). Childhood sex-typed behavior and sexual orientation. *Dev. Psychol.* 31, 43–55. doi: 10.1037/0012-1649.31.1.43

Banse, R., Seise, J., and Zerbes, N. (2001). Implicit attitudes towards homosexuality: reliability, validity, and controllability of the IAT. *Zeitschrift Exp. Psychol.* 48, 145–160. doi: 10.1026//0949-3946.48.2.145

Beere, C. A. (1990). *Gender Roles: A Handbook of Tests and Measures.* Westport, CT: Greenwood Press.

Bem, S. L. (1974). The measurement of psychological androgyny. *J. Consult. Clin. Psychol.* 42, 155–162. doi: 10.1037/h0036215

Bem, S. L. (1979). Theory and measurement of androgynity: a reply to the pedhazur-tetenbaum and locksley-colten critiques. *J. Pers. Soc. Psychol.* 37, 1047–1054. doi: 10.1037/0022-3514.37.6.1047

Blashill, A. J., and Powlishta, K. K. (2009). Gay stereotypes: the use of sexual orientation as a cue for gender-related attributes. *Sex Roles* 61, 783–793. doi: 10.1007/s11199-009-9684-7

Bortz, J., and Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler [Research Methods and Evaluation for Human and Social Scientists]*. Heidelberg: Springer Medizin Verlag.

Bühner, M. (2010). *Einführung in Die Test- und Fragebogenkonstruktion [Introduction to Test and Questionnaire Construction*]. München: Pearson Studium.

Camp, M. (2009). *Japanese Lesbian Speech: Sexuality, Gender Identity, and Language*. Doctoral dissertation, University of Arizona. Available online at: http://arizona.openrepository.com/arizona/bitstream/10150/195371/1/azu_etd_10574_sip1_m.pdf

Choi, N., and Fuqua, D. R. (2003). The structure of the bem sex role inventory: a summary report of 23 validation studies. *Educ. Psychol. Meas.* 63, 872–887. doi: 10.1177/0013164403258235

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Academic Press.

Coleman, E. (1987). Assessment of sexual orientation. *J. Homosex.* 14, 9–24. doi: 10.1300/J082v14n01_02

Constantinople, A. (1973). Masculinity-femininity: an exception to a famous dictum? *Psychol. Bull.* 80, 389–407. doi: 10.1037/h0035334

Croft, A., Schmader, T., Block, K., and Baron, A. S. (2014). The second shift reflected in the second generation: Do parents' gender roles at home predict children's aspirations? *Psychol. Sci.* 25, 1–11. doi: 10.1177/0956797614533968

Deaux, K. (1984). From individual differences to social categories: analysis of a decade's research on gender. *Am. Psychol.* 39, 105–116. doi: 10.1037/0003-066X.39.2.105

Deaux, K., and Lewis, L. L. (1984). Structure of gender stereotypes: Interrelationships among components and gender label. *J. Pers. Soc. Psychol.* 46, 991–1004. doi: 10.1037/0022-3514.46.5.991

Diekman, A. B., and Eagly, A. H. (2000). Stereotypes as dynamic constructs: women and men of the past, present, and future. *Pers. Soc. Psychol. Bull.* 26, 1171–1188. doi: 10.1177/0146167200262001

Ebert, I. D., Steffens, M. C., and Kroth, A. (2014). Warm, but maybe not so competent? – Contemporary implicit stereotypes of women and men in Germany. *Sex Roles* 70, 359–375. doi: 10.1007/s11199-014-0369-5

Evers, A., and Sieverding, M. (2014). Why do highly qualified women (still) earn less? Gender differences in long-term predictors of career success. *Psychol. Women Q.* 38, 93–106. doi: 10.1177/0361684313498071

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146

Fiske, S. T., Cuddy, A. J. C., Glick, P., and Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *J. Pers. Soc. Psychol.* 82, 878–902. doi: 10.1037/0022-3514.82.6.878

Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* 74, 1464–1480. doi: 10.1037/0022-3514.74.6.1464

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis. A Regression-Based Approach*. New York, NY: Guilford Press.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., and Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Pers. Soc. Psychol. Bull.* 31, 1369–1385. doi: 10.1177/0146167205275613

Howitt, D., and Cramer, D. (2008). *Introduction to Research Methods in Psychology*. Essex: Pearson Education.

Kagan, J. (1964). "Acquisition and significance of sex-typing and sex-role identity," in *Review of Child Development Research*, eds M. L. Hoffman and L. W. Hoffman (New York, NY: Russell Sage), 17–43.

Kite, M. E., and Deaux, K. (1987). Gender belief systems: homosexuality and the implicit inversion. *Psychol. Women Q.* 11, 83–96. doi: 10.1111/j.1471-6402.1987.tb00776.x

Kite, M. E., and Whitley, B. E. Jr. (1996). Sex differences in attitudes toward homosexual persons, behaviors, and civil rights: A meta-analysis. *Pers. Soc. Psychol. Bull.* 22, 336–353. doi: 10.1177/0146167296224002

Leach, C., Van Zomeren, M., Zebel, S., Vliek, M. W., Pennekamp, S. F., Doosje, B., et al. (2008). Group-level self-definition and self-investment: A hierarchical (multicomponent) model of in-group identification. *J. Pers. Soc. Psychol.* 95, 144–165. doi: 10.1037/0022-3514.95.1.144

Lippa, R. A. (2005). Sexual orientation and personality. *Annu. Rev. Sex Res.* 16, 119–153. doi: 10.1080/10532528.2005.10559831

Lippa, R. A. (2008). The relation between childhood gender nonconformity and adult masculinity–femininity and anxiety in heterosexual and homosexual men and women. *Sex Roles* 59, 684–693. doi: 10.1007/s11199-008-9476-5

Madon, S. (1997). What do people believe about gay males? A study of stereotype content and strength. *Sex Roles* 37, 663–685. doi: 10.1007/BF02936334

Munson, B., and Babel, M. (2007). Loose lips and silver tongues, or, projecting sexual orientation through speech. *Lang. Linguist. Compass* 1, 416–449. doi: 10.1111/j.1749-818X.2007.00028.x

Niedlich, C., Steffens, M. C., Krause, J., Settke, E., and Ebert, I. D. (2015). Ironic effects of sexual minority group membership: Are lesbians less susceptible to invoking negative female stereotypes than heterosexual women? *Arch. Sex. Behav.* 44, 1439–1447. doi: 10.1007/s10508-014-0412-1

Nosek, B. A., Banaji, M. R., and Greenwald, A. G. (2002). Math = male, me = female, therefore math ≠ me. *J. Pers. Soc. Psychol.* 83, 44–59. doi: 10.1037/0022-3514.83.1.44

Nosek, B. A., Greenwald, A. G., and Banaji, M. R. (2005). Understanding and using the implicit association Test: II. Method variables and construct validity. *Pers. Soc. Psychol. Bull.* 31, 166–180. doi: 10.1177/0146167204271418

Pierrehumbert, J. B., Bent, T., Munson, B., Bradlow, A., and Bailey, J. M. (2004). The influence of sexual orientation on vowel production. *J. Acoust. Soc. Am.* 116, 1905–1908. doi: 10.1121/1.1788729

Rieger, G., Linsenmeier, J. A. W., Gygax, L., Garcia, S., and Bailey, J. M. (2010). Dissecting "gaydar": accuracy and the role of masculinity–femininity. *Arch. Sex. Behav.* 39, 124–140. doi: 10.1007/s10508-008-9405-2

Roth, J., and Mazziotta, A. (2015). Adaptation and validation of a German multidimensional and multicomponent measure of social identification. *Soc. Psychol.* 46, 277–290. doi: 10.1027/1864-9335/a000243

Runge, T. E., Frey, D., Gollwitzer, P. E., Helmreich, R. L., and Spence, J. T. (1981). Masculine (instrumental) and feminine (expressive) traits: a comparison between students in the United States and West Germany. *J. Cross Cult. Psychol.* 12, 142–162. doi: 10.1177/0022022181122002

Safir, M. P., Rosenmann, A., and Kloner, O. (2003). Tomboyism, sexual orientation, and adult gender roles among Israeli women. *Sex Roles* 48, 401–410. doi: 10.1023/A:1023522410882

Sánchez, F., and Vilain, E. (2012). "Straight-Acting Gays": The relationship between masculine consciousness, anti-effeminacy, and negative gay identity. *Arch. Sex. Behav.* 41, 111–119. doi: 10.1007/s10508-012-9912-z

Sczesny, S., Bosak, J., Neff, D., and Schyns, B. (2004). Gender stereotypes and the attribution of leadership traits: A cross-cultural comparison. *Sex Roles* 51, 631–645. doi: 10.1007/s11199-004-0715-0

Sidanius, J., Pratto, F., and Bobo, L. (1994). Social dominance orientation and the political psychology of gender: A case of invariance? *J. Pers. Soc. Psychol.* 67, 998–1011. doi: 10.1037/0022-3514.67.6.998

Smiler, A. P. (2004). Thrity years after the discovery of gender: psychological concepts and measures of masculinity. *Sex Roles* 50, 15–26. doi: 10.1023/B:SERS.0000011069.02279.4c

Spence, J. T., and Buckner, C. E. (2000). Instrumental and expressive traits, trait stereotypes, and sexist attitudes. *Psychol. Women Q.* 24, 44–62. doi: 10.1111/j.1471-6402.2000.tb01021.x

Spence, J. T., Helmreich, R. L., and Stapp, J. (1975). Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity. *J. Pers. Soc. Psychol.* 32, 29–39. doi: 10.1037/h0076857

Spence, J. T., Helmreich, R. L., and Stapp, J. (1978). *Masculinity and femininity: Their Psychological Dimensions, Correlations, and Antecedents.* Austin, TX: University of Texas Press.

Steffens, M. C., and Buchner, A. (2003). Implicit association test: Separating transsituationally stable and variable components of attitudes toward gay men. *Exp. Psychol.* 50, 33–48. doi: 10.1027//1618-3169.50.1.33

Steffens, M. C., Jelenec, P., and Noack, P. (2010). On the leaky math pipeline: Comparing implicit math-gender stereotypes and math withdrawal in female and male children and adolescents. *J. Educ. Psychol.* 102, 947–963. doi: 10.1037/a0019920

Steffens, M. C., Kirschbaum, M., and Glados, P. (2008). Avoiding stimulus confounds in Implicit Association Tests by using the concepts as stimuli. *Br. J. Soc. Psychol.* 47, 217–243. doi: 10.1348/014466607X226998

Steffens, M. C., and Schulze-Koenig, S. (2006). Predicting spontaneous Big Five behavior with Implicit Association Tests. *Eur. J. Psychol. Assess.* 22, 13–20. doi: 10.1027/1015-5759.22.1.13

Steffens, M. C., and Viladot, M. A. (2015). *Gender at Work: A Social-Psychological Perspective*. New York, NY: Peter Lang.

Strack, F., and Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Pers. Soc. Psychol. Rev.* 8, 220–247. doi: 10.1207/s15327957pspr0803_1

Swank, E., and Raiz, L. (2010). Attitudes toward gays and lesbians among undergraduate social work students. *J. Women Soc. Work* 25, 19–29. doi: 10.1177/0886109909356058

Swazina, K. R., Waldherr, K., and Maier, K. (2004). Geschlechtsspezifische Ideale im Wandel der Zeit [Gender-specific ideals through the ages]. *Zeitschrift Differentielle Diagnostische Psychologie* 25, 165–176. doi: 10.1024/0170-1789.25.3.165

Thompson, E. H. Jr., and Bennet, K. M. (2015). Measurement of masculinity ideologies: a (Critical) review. *Psychol. Men Masculinity* 16, 115–133. doi: 10.1037/a0038609

Valentova, J., Rieger, G., Havlicek, J., Linsenmeier, J. A., and Bailey, J. M. (2011). Judgments of sexual orientation and masculinity-femininity based on thin slices of behavior: A cross-cultural comparison. *Arch. Sex. Behav.* 40, 1145–1152. doi: 10.1007/s10508-011-9818-1

Wilde, A., and Diekman, A. B. (2005). Cross-cultural similarities and differences in dynamic stereotypes: A comparison between Germany and the United States. *Psychol. Women Q.* 29, 188–196. doi: 10.1111/j.1471-6402.2005.00181.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Manuscript 1 – Related Appendix

Appendix A

A1. The Traditional Masculinity-Femininity Scale (TMF) in English translation [with original German wording]

- I consider myself as…
  [Ich empfinde mich selbst als…]
- Ideally, I would like to be…
  [Idealerweise wäre ich gern…]
- Traditionally, my interests would be considered as…
  [Traditionellerweise würden meine Interessen angesehen werden als…]
- Traditionally, my attitudes and beliefs would be considered as…
  [Traditionellerweise würden meine Einstellungen und Ansichten angesehen werden als…]
- Traditionally, my behavior would be considered as…
  [Traditionellerweise würde mein Verhalten angesehen werden als…]
- Traditionally, my outer appearance would be considered as…
  [Traditionellerweise würde meine äußere Erscheinung angesehen werden als…]

Scales ranged from 1 (*not at all masculine*) to 7 (*totally masculine*) and from 1 (*not at all feminine*) to 7 (*totally feminine*) in Study 1 and from 1 (*totally masculine*) to 7 (*totally feminine*) in the pilot study and in Studies 2-3.

A2. Childhood Gender-Role Behavior Scale (CGRB)

1. In my childhood I liked being a girl.
2. In my childhood I would rather have been a boy.
3. In my childhood I behaved like a “typical girl”.
4. In my childhood I behaved like a “typical boy”.
5. In my childhood I played typical girl games.
6. In my childhood I played typical boy games.
7. In my childhood I played with girls.
8. In my childhood I played with boys.
9. In my childhood I wanted to wear dresses and skirts.
10. In my childhood I wanted to wear trousers.

Scales ranged from 1 (*never applied*) to 7 (*always applied*) in Study 1 and from 1 (*I strongly disagree*) to 5 (*I strongly agree*) in Study 2.

Appendix B

Table 1 Factor, pattern, and structure matrix of the exploratory factor analysis with a two-factor solution and the confirmatory one-factor analysis of the TMF in Study 1.

		Factor Matrix		Pattern Matrix		Structure Matrix		Factor matrix CFA
		F1	F2	F1	F2	F1	F2	
1.	(F) I consider myself as…	.71	.39	-.09	.87	.51	.81	.70
•	(F) Ideally, I would like to be...	.77	.43	-.11	.95	.55	.88	.75
•	(F) Traditionally, my interests would be considered as…	.76	-.27	.76	.06	.80	.58	.75
3.	(F) Traditionally, my attitudes and beliefs would be considered as…	.78	-.37	.90	-.06	.86	.56	.77
4.	(F) Traditionally, my behavior would be considered as…	.87	-.22	.77	.18	.89	.70	.87
5.	(F) Traditionally, my outer appearance would be considered as…	.77	.27	.10	.74	.61	.81	.77
6.	(M) I consider myself as…	-.80	-.21	-.19	-.69	-.66	-.82	-.80
7.	(M) Ideally, I would like to be...	-.67	-.23	-.09	-.64	-.53	-.70	-.67
8.	(M) Traditionally, my interests would be considered as…	-.81	.27	-.81	-.07	-.85	-.62	-.81
9.	(M) Traditionally, my attitudes and beliefs would be considered as…	-.69	.36	-.84	.10	-.76	-.48	-.68
10.	(M) Traditionally, my behavior would be considered as…	-.79	.22	-.73	-.13	-.82	-.63	-.79
11.	(M) Traditionally, my outer appearance would be considered as…	-.75	-.25	-.11	-.71	-.60	-.78	-.75

Note. Scale ranged from 1 - "not all masculine/feminine" to 7 - "very masculine/feminine". Abbreviations: Factor 1 (F1), Factor 2 (F2), confirmatory factor analysis (CFA), femininity scale (F), masculinity scale (M)

Manuscript 2

Kachel, S., Simpson, A. P., & Steffens, M. C. (2017).
Acoustic Correlates of Sexual Orientation and Gender-Role Self-Concept in Women's Speech.
Journal of the Acoustical Society of America, 141, 4793–4809.
DOI*:* 10.1121/1.4988684

PROOF COPY [JASA-00586] 066706JAS

Acoustic correlates of sexual orientation and gender-role self-concept in women's speech

Sven Kachel[a)]
Department of Social, Environmental, and Economic Psychology, University of Koblenz-Landau, Forststraße 7, 76829 Landau, Germany

Adrian P. Simpson
Department of German Linguistics, Friedrich Schiller University Jena, Fürstengraben 30, 07743 Jena, Germany

Melanie C. Steffens
Department of Social, Environmental, and Economic Psychology, University of Koblenz-Landau, Forststraße 7, 76829 Landau, Germany

(Received 18 March 2016; revised 13 May 2017; accepted 7 June 2017; published online xx xx xxxx)

Compared to studies of male speakers, relatively few studies have investigated acoustic correlates of sexual orientation in women. The present investigation focuses on shedding more light on intra-group variability in lesbians and straight women by using a fine-grained analysis of sexual orientation and collecting data on psychological characteristics (e.g., gender-role self-concept). For a large-scale women's sample (overall $n = 108$), recordings of spontaneous and read speech were analyzed for median fundamental frequency and acoustic vowel space features. Two studies showed no acoustic differences between lesbians and straight women, but there was evidence of acoustic differences *within* sexual orientation groups. Intra-group variability in median f0 was found to depend on the exclusivity of sexual orientation; F1 and F2 in /iː/ (study 1) and median f0 (study 2) were acoustic correlates of gender-role self-concept, at least for lesbians. Other psychological characteristics (e.g., sexual orientation of female friends) were also reflected in lesbians' speech. Findings suggest that acoustic features indexicalizing sexual orientation can only be successfully interpreted in combination with a fine-grained analysis of psychological characteristics.
 [http://dx.doi.org/10.1121/1.4988684]

[LK] Pages: 1–17

I. INTRODUCTION

"But how does a lesbian speak?" or "I would prefer to do a gay man." were frequent responses of speakers in our laboratory when they were instructed to imitate lesbian speaking patterns. Knowledge of stereotypes regarding lesbian speech seems to be weak (but see Queen, 1997). This asymmetry is reflected in the fact that studies referring to phonetic correlates of sexual orientation have largely concentrated on differences between gay and straight male speakers (for a review, see Munson and Babel, 2007). By comparison, relatively few studies have compared acoustic features of the speech of self-identified lesbians and straight women. In addition, these studies' findings are inconsistent. The aim of the present research is to shed more light on differences in women's speech depending on actual sexual orientation by taking psychological characteristics into account.

A. Phonetic research on sexual orientation

Most phonetic studies in the field of sexual orientation and speech have investigated whether lesbians (respectively, gay men) would mirror speaking patterns of straight men (respectively, straight women). This assumption is based on lay people's stereotype-driven gender inversion theory (Kite and Deaux, 1987, henceforth: gender inversion theory). Studies dealing with the effect of female sexual orientation on speaking patterns have primarily concentrated on intonational features and vowel characteristics. Findings regarding intonational features are inconsistent. This is particularly true for mean fundamental frequency (f0). The first investigation (Moonwomon-Baird, 1997) showed that a dyad of straight women used a higher mean f0 than a lesbian dyad when having a conversation (spontaneous speech). However, this study can only be used as anecdotal evidence due to methodological shortcomings (see Jacobs, 1996; Waksler, 2001). Nevertheless, the finding is supported by a study on Japanese women investigating spontaneous speech during an interview (Camp, 2011). On average, 7 straight women spoke two semi-tones higher than 10 non-straight women (7 lesbian, 3 bisexual). van Borsel *et al.* (2013) replicated that finding for a Dutch female sample: 68 straight women spoke on average 10 Hz higher than an age-matched group of 34 self-identified lesbians when reading a short story. Whereas the three studies focusing on connected speech converged on the finding that straight women speak with a higher mean f0 than lesbians and thus provided scientific evidence for gender inversion theory, two other studies concerning

[a)]Electronic mail: sven.kachel@uni-jena.de

vowel-specific f0 values from gender-mixed, sexually divergent samples failed to find a significant effect of sexual orientation on mean f0 in women. Neither Munson *et al.* (2006), who compared 11 non-straight and straight women from Minneapolis when reading a list of single words, nor Rendall *et al.* (2008), who contrasted 29 non-straight (10 bisexual) and 33 straight women from Alberta (Canada) when reading four sentences and single words, provided any evidence for a significant influence of sexual orientation on f0.

Similarly, findings for vowel characteristics are also inconclusive. Pierrehumbert *et al.* (2004) analyzed first and second formant frequency values of the five target vowels /ɑ i eɪ u æ/ from a sentence list in recordings of 16 straight and 32 non-straight (16 bisexual) Chicago women. Straight women showed significantly higher mean F1 and F2 values that were due to differences in back vowels (F1 and F2 in /u/ and F1 in /ɑ/). In other words, straight women produced their vowel space at a greater distance from the male vowel space area than non-straight women, which is in line with gender inversion theory. However, in the same study, straight women unexpectedly exhibited a more contracted vowel space and thus were acoustically closer to straight men than lesbians, which is not in line with gender inversion theory. Whereas Munson *et al.* (2006) completely failed to replicate these differences in vowel space shift and expansion using the 8 vowels /i ɪ eɪ ɛ æ ɔɪ oʊ u/ derived from 32 consonant-vowel-consonant words, Rendall *et al.* (2008) did provide supporting evidence for vowel space shift: straight women showed significantly higher mean formant frequencies than lesbians (ΔF1: 25 Hz; ΔF2: 41 Hz). However, in contrast to Pierrehumbert *et al.* (2004), Rendall *et al.* (2008) descriptively found the opposite result pattern for vowel space expansion: Straight women's vowel spaces tended to be more expanded than those of lesbians.

In comparison to aggregated vowel space characteristics, the pattern of differences in specific formant values of single vowels is even more inconsistent. Seven common vowels (/i eɪ ɪ ɛ æ u oʊ/) were used in at least two of the investigations that measured vowel characteristics (Pierrehumbert *et al.*, 2004; Munson *et al.*, 2006; Rendall *et al.*, 2008). Overall, the three studies showed nine differences between lesbians and straight women in F1 or F2, but only one replication: Munson *et al.* (2006) and Rendall *et al.* (2008) showed straight women to produce /ɛ/-sounds with higher F1- and F2-values on average, making them acoustically more distinct from straight men than lesbians. All other possible replications were not obtained. This has been interpreted as evidence in support of diversity within lesbian and straight speakers using a myriad of speaking patterns which consequently makes a final detection of all phonetic markers an almost impossible undertaking (see Zwicky, 1997). Alternatively, failure to replicate may be due to shortcomings of previous studies.

In short, some acoustic markers of sexual orientation in women have been identified, but have not been consistently replicated. A number of reasons for this are conceivable. First, data have been elicited from speakers carrying out a number of different linguistic activities in different communicative situations, ranging from naturally occurring talk (Moonwomon-Baird, 1997) to the utterance of single monosyllabic words (Munson *et al.*, 2006). Second, previous studies have often investigated relatively small numbers of lesbians ($\leq$16, except van Borsel *et al.*, 2013; Rendall *et al.*, 2008). Third, a coarse grouping of speakers was undertaken, treating lesbians and bisexual women as a single category. This is problematic for two reasons. First, it suggests a clear intergroup demarcation between straight and non-straight speakers, and second, it suggests homogeneity within the groups (see also Levon, 2006). Strictly speaking, no previous study, except van Borsel *et al.* (2013), has investigated which acoustic features are used by lesbians. Lumping together lesbians and bisexual women seems inappropriate for at least two reasons. By referring directly to sexual orientation, bisexual women were long excluded from and by the lesbian community because of their attraction to men. Not to be involved with men was a central criterion for defining what it means to be a lesbian (e.g., Jeffreys, 1999; Walters, 1996). By referring to further psychological differences, bisexual women were found to be less masculine/more feminine than lesbians but more masculine/less feminine than straight women (e.g., Lippa, 2005). Hence, it seems reasonable that when we study possible correlations between sexual orientation and speech characteristics we should initially separate groups using more fine-grained measures of sexual orientation.

As a final reason for inconsistent findings, studies of the phonetic correlates of sexual orientation have largely failed to include related psychological characteristics that possibly explain within-group variance (see also Waksler, 2001; Munson and Babel, 2007). Decades of sociolinguistic research have shown that speakers can accommodate to a multiplicity of axes of social differentiation, such as nationality, age, and social class. There have been a number of attempts from an intersectional perspective to illuminate within-category variability by interdigitating gender with other social categories, like social class (Stuart-Smith, 2007) or sexual orientation. However, biophysical and social characteristics of gender cannot be treated as one thing (see Eckert, 1990) and their effects on speech should be separated (see Smyth and Rogers, 2008). Thus, variability in speech characteristics within groups of lesbians and straight women could be due to other variables than sexual orientation.

B. Psychological characteristics related to sexual orientation

Researchers who have conducted sociophonetic studies considering women's sexual orientation have speculated on psychological characteristics that could mediate or moderate associations between sexual orientation and acoustic parameters. First, it has been suggested that *group affiliation*, rather than sexual orientation, influences speech (Munson *et al.*, 2006; Munson, 2011; Waksler, 2001). Because people may identify as lesbian, for instance, but reject affiliation to the lesbian community (Szymanski and Chung, 2001), the present study adopted two different measures for *sexual* *orientation* and *group affiliation*, which is defined as the degree to which a person sees herself as a group member. Additionally, we measured *psycho-social identification* with the in-group: the importance of and satisfaction with one's group membership. We tested whether lesbians who had a high affiliation to straight women and a low affiliation to lesbians, and who displayed a low psycho-social identification with the in-group, showed speaking patterns consistent with stereotypical straight women's speech.

A second psychological intragroup difference could be *present gender-role self-concept* (Pierrehumbert *et al.*, 2004; Munson and Babel, 2007; Rendall *et al.*, 2008), defined as the extent to which a person describes herself on a rather global level as typically masculine and/or feminine (see Kachel *et al.*, 2016). Straight women often describe themselves as more feminine and less masculine than lesbians on gender-related scales according to stereotypical gender-role self-concepts (e.g., Lippa, 2005; Spence and Helmreich, 1978; Kachel *et al.*, 2016) supporting gender-inversion theory (Kite and Deaux, 1987). Also, there is considerable variance in present gender-role self-concepts among straight women (Kachel *et al.*, 2016) and among lesbians who could be classified as either more masculine (so-called butch-type) or feminine (so-called femme-type) (Loulan, 1990; Pearcey *et al.*, 1996; also see Singh *et al.*, 1999). Moreover, some researchers have assumed a phonetic acquisition scenario by referring to childhood as the critical phase (Smyth and Rogers, 2008; van Borsel *et al.*, 2009). Indeed, regarding *gender-role self-concept during childhood*, straight women were found to remember more gender-typical behavior than lesbians (Bailey and Zucker, 1995). We therefore tested whether speech differences between lesbians and straight women were due to differences in gender-role self-concept during childhood or adulthood (i.e., at present). We further assumed that lesbians and straight women who conformed to a more feminine gender role would produce commensurate speaking patterns.

A third group of psychological characteristics that could mediate or moderate associations between sexual orientation and acoustic correlates is formed by different *social environments*. Similarly to children imitating their peers' behavior, adolescent or adult lesbians could direct their attention towards lesbian role-models (Smyth and Rogers, 2008; Pierrehumbert *et al.*, 2004). Therefore, we collected data regarding women's *childhood* (contact to boys and girls) and *present social environment* (e.g., sexual orientation of female friends). We tested whether lesbians who reported typically feminine social environment characteristics produce speaking patterns in line with feminine gender role.

Finally, psychological intragroup differences could be related to lesbians' *coming-out measures*. On the one hand, phonetic markers stereotypically associated with lesbian speech should only be found for lesbians who had come out as lesbians and who are comfortable with their lesbian identity (van Borsel *et al.*, 2009). Those who have come out should not try to conceal any acoustic cues associated with the lesbian group. Whereas lesbians' coming-out age had no effect on mean fundamental frequency or variation of fundamental frequency when age was held constant (van Borsel *et al.*, 2013; for men please see Baeck *et al.*, 2011), lesbians could try to hide their sexual orientation because of a high perceived risk of discrimination (e.g., Steffens and Wagner, 2004) or because of negative coming-out experiences. Consequently, those lesbians who report less discrimination, fewer negative coming-out experiences, and a higher coming-out extent can be expected to produce speaking patterns that are more typical of lesbians.

C. The present research

In the present investigation we attempted to remedy some of the weaknesses we identified in earlier studies. In line with previous investigations, we analyzed typical lab speech, drawn from spoken prose and spontaneous speech. Our sample sizes comprised in total over 100 female speakers, split evenly across groups of lesbian and straight speakers. The most important aspects of the present studies were the greater intragroup differentiation with respect to sexual orientation as well as the analysis of a range of psychological characteristics.

We are asking the following questions: What are the acoustic correlates of sexual orientation in German-speaking women? Are there significant acoustic differences between lesbians and straight women? Does intragroup homogeneity exist within the straight and lesbian speaker groups, or are there correlations between acoustic parameters and sexual self-identification *within* the lesbian and straight groups? Do psychological characteristics influence lesbians' speech? And finally, are there acoustic correlates of gender-role self-concept?

To answer these questions two studies—that have been part of a larger research project—were carried out with female speakers. Prior to recording, speakers completed an online questionnaire to collect information about sexual orientation and related psychological characteristics. The questionnaire ensured that the sample contained the same number of lesbian and straight speakers. We examined median fundamental frequency, vowel space shift and dispersion, as well as formant frequencies of single vowels, and vowel durations. The first study was carried out in Jena, a German city with approximately 100 000 inhabitants; the second study, designed to replicate and extend the first, was carried out in the German capital, Berlin. Regarding psychological characteristics, we tested whether gender-role self-concept has its own effects on women's speech and whether lesbians' speech is influenced by any of the psychological characteristics.

II. STUDY 1

The aim of the first study was to investigate the relationships between acoustic representations of sexual orientation and gender by considering psychological characteristics to shed light on group variability. For a detailed overview about all items of the psychological measures, please see the supplemental information.[1]

A. Method

1. Participants

Fifty-one female speakers living in east central Germany were invited to take part in an investigation on "Sexuality and voice" via bulletin boards, handouts, and mailing lists at the University of Jena, as well as directly through Lesbian, gay, bisexual, transgender, intersexual, and queer (LGBTIQ*) organizations and platforms. Using a snowball technique, speakers were asked to encourage friends and acquaintances to participate as well. In order to reach an adequate number of speakers per group, the study description explicitly contained a request for women divergent in sexual orientation. In addition, it was mentioned that only German native speakers without voice and speaking disorders could be considered.

The sample was homogeneous with regard to ethnicity, educational level, and age. Nearly all women were students, were Caucasian, and ranged in age from 20 to 30 years [mean $(M) = 23.37$, standard deviation $(SD) = 2.40$]. According to the 7-point Kinsey-like Scale that we used as a subjective measure of sexual orientation (see Kinsey *et al.*, 1953), 22 women identified as lesbian (Kinsey-like scores ≤ 3) and 22 women identified as straight (Kinsey-like scores ≥ 5). This group assignment allows us to check for group variability within sexual orientations. Hence, seven bisexual women (Kinsey-like score 4) were excluded from analyses except for comparisons of non-straight (Kinsey-like scores 1–5) and straight women (Kinsey-like scores 6–7) when trying to replicate previous findings which were based on the comparison of non-straight and straight women (detailed distributions and comparison of classification strategies can be seen in Table I). Lesbians and straight women did not differ significantly in terms of sociodemographic characteristics such as age ($M_{\text{lesbian}} = 23.91$, $M_{\text{straight}} = 22.86$; $t[42] = 1.46$, $p = 0.15$).

2. Psychological measures

Following initial recruitment, speakers were asked to complete an online questionnaire eliciting information about sexual orientation as well as a range of related psychological attributes. All internal consistencies of the described scales were Cronbach's $\alpha \geq 0.76$, indicating reliable measurement.

a. Sexual orientation. Because there is an on-going debate about whether objectifiable measures, such as actual sexual behavior, are more valid in determining sexual orientation (e.g., Barrett, 1997; Kulick, 2000), in addition to sexual self-identification, we used a more objectifiable instrument to measure sexual orientation (Worthington and Reynolds, 2009; also see Shively *et al.*, 1985; Klein *et al.*, 1985). Sexual orientation towards women and men was measured separately with four items each: sexual fantasies, romantic emotions, physical attraction, and sexual interaction (e.g., "Physical attraction towards women"). These were rated in order to answer the question "How often have you experienced" on a 7-point scale ("1 = *never*", "7 = *always*"). Higher scores indicated higher degrees of sexual orientation towards women/men. Sexual orientation towards women and men intercorrelated ($r = -0.75$, $p < 0.001$). Additionally, correlations between the Kinsey-like Scale and objectifiable sexual orientation were high ($r_{\text{towards women}} = -0.96$, $p < 0.001$; $r_{\text{towards men}} = 0.83$, $p < 0.001$).

b. Present gender-role self-concept. Two questionnaires were used for measuring present gender-role self-concept. First, speakers rated themselves on the 7-point ("1 = *very masculine*," "7 = *very feminine*") Traditional Masculinity-Femininity Scale (TMF; Kachel *et al.*, 2016), which consists of six bipolar items such as "I consider myself…" and "Traditionally, my behavior would be regarded as…." For each person an individual scale mean was calculated, with higher scores indicating more femininity/less masculinity.

Second, speakers completed the German version (GEPAQ; Runge *et al.*, 1981) of the Extended Personal Attributes Questionnaire (Spence and Helmreich, 1978). This is a more specific measure of present gender-role self-concept, given that it independently collects information for masculinity and femininity associated with personality characteristics without referring directly to gender. In order to separate masculinity from femininity, we used two scales with eight items each. On the first scale (GEPAQ-M), instrumental, goal-oriented traits, typically associated with men, like independence and self-confidence, were assessed. On the second scale (GEPAQ-F), stereotypical female traits which were more social-relations oriented and related to emotional expressiveness, such as helpfulness and kindness, were assessed on a 5-point scale. Two mean values for each person were computed with higher scores indicating a higher

TABLE I. Distribution of speakers regarding sexual orientation on Kinsey-like Scale in studies 1 and 2 and comparison of two sexual orientation classification strategies.

Kinsey-like score		Number per cell		Classification strategy	
		Study 1	Study 2	Lesbians vs straight	Non-straight vs straight
1	Exclusively lesbian	7	9	Lesbians	Non-straight women
2	Mainly lesbian	12	15	Lesbians	Non-straight women
3	Somewhat lesbian	3	3	Lesbians	Non-straight women
4	Equally lesbian and straight	7	3		Non-straight women
5	Somewhat straight	2	6	Straight women	Non-straight women
6	Mainly straight	8	12	Straight women	Straight women
7	Exclusively straight	12	9	Straight women	Straight women

degree of masculinity on GEPAQ-M and a higher degree of femininity on GEPAQ-F.

c. Group affiliation and psycho-social identification. Two circle items (see Schubert and Otten, 2002) were used to measure the degree of self-assigned affiliation to the groups of lesbians and straight women. Each item consisted of seven graphics visualizing the distance between a smaller circle (self) and a larger circle (group). The smaller the distance and the higher the overlap between the two circles, the higher the degree of group affiliation (see supplemental material for visualization).

In addition to this pictorial measurement, two word-based scales were taken from Luhtanen and Crocker's (1992) Collective Self-Esteem Scale in order to assess speakers' psycho-social identification as lesbian or straight. Both measures contained four items that were rated on a 7-point scale ("1 = *strong rejection*," "7 = *strong agreement*"). First, the "Identity"-subscale was used to determine the group membership's importance (e.g., "The social group I belong to is an important reflection of who I am"). Second, the "Private"-subscale was applied to assess speakers' satisfaction with group membership (e.g., "In general, I'm glad to be a member of the social group I belong to"). In reliability analyses, lower internal consistencies for both subscales (Cronbach's $\alpha \leq 0.70$) compared to the overall scale were discovered (Cronbach's $\alpha = 0.80$). Thus, an overall index of psycho-social identification with the in-group was computed. Higher scores indicated a higher psycho-social identification with the in-group.

d. Present social environment. On the basis of Hodson *et al.* (2009) speakers evaluated their current contact to lesbians and straight women on two separate 7-point scales ("1 = *never*," "7 = *always*"). Data for general gender distribution of friends and sexual orientation of female friends were also collected on two 7-point scales ("1 = *men/lesbians only*," "7 = *women/straight women only*"). Higher scores indicated a higher proportion of female and straight female friends.

e. Coming-out measures. Two measures related to speakers' coming-out were collected from non-straight women only. The coming-out extent represents the visibility of same-gender orientation to the social environment as well as negative experiences during the coming-out process. Coming-out extent was assessed using a questionnaire partly based on Herek *et al.* (1997). Speakers rated parental knowledge of their non-straight orientation for mother and father separately. Subsequently, they provided information on how many persons of different groups (present straight friends, acquaintances, fellow students, colleagues, and boss) know about their sexual orientation ("1 = *none*," "7 = *all of them*"). In order to obtain comparable values for the parents' questions, the scores were adapted to the 7-point scale: If neither mother nor father knew about the sexual orientation, this was recoded as "1"; if only one parent knew about it, it was represented with "4", and if both parents knew, "7" was assigned. The mean of all items was computed with higher scores indicating a higher coming-out extent.

Three subscales of the Lesbian and Gay Identity Scale (Mohr and Kendra, 2011) were used to examine negative coming-out experiences. Each scale consisted of three items to be rated on a 6-point scale ("1 = *strong rejection*," "6 = *strong agreement*"). Problems regarding coming-out were assessed with the Difficult Process Scale (e.g., "Admitting to myself that I'm an LGB person has been a very painful process"). The Concealment Motivation Scale with items like "I keep careful control over who knows about my same-gender romantic relationships" assessed the speakers' preference to hide their sexual orientation. The Internalized Homonegativity Scale gathered data on speakers' wishes to change their sexual orientation, including items like "If it were possible, I would choose to be straight." Means were computed for all subscales with higher scores indicating more negative coming-out experiences.

3. Procedure and equipment

Data collection comprised two phases. In the first phase, 75 women completed an online questionnaire which was designed to obtain a sample evenly distributed across lesbian and straight speakers. In the second, *in situ* investigation phase, voices of a subset of the speakers ($n = 51$) were recorded in a sound-treated room; 24 out of the 75 women who initially took part were not invited to the second phase, because they self-identified as straight (Kinsey-like scores ≥ 5). All recordings were made by the same male, Caucasian, mid-20 year-old experimenter wearing a dark shirt and dark pants. Recordings were done using an AKG C1000S capacitor microphone connected to a preamplifying audio interface (M-AUDIO Fast Track Pro) at a sampling rate of 44.1 kHz and 16-bit amplitude resolution.

Each speaker recorded samples of read and spontaneous speech. The acoustic analyses reported in both studies used data from a particular sentence list (Weirich and Simpson, 2013) that was directly read from paper three times. Included in a list of 20 sentences were three sentences shown in (1), containing the abbreviations *IAA, BII, LUU* which were designed to elicit sequences of long point vowel tokens /aː iː uː/, minimally affected by any neighboring consonants or temporal constraints giving rise to undershoot. In most varieties of German, including the Standard, we would also expect speakers to exhibit a certain amount of junctural glottalization between the two vowels, often taking the form of creaky voice (Kohler, 1994). The low fundamental frequency typically associated with different types of glottalization ensured tight interharmonic spacing, in turn allowing for more reliable formant estimation, especially in otherwise higher pitched female voices (Weirich and Simpson, 2013).

(1) Sie fuhren letzte Woche zur IAA nach Frankfurt
("They went to the IAA in Frankfurt last week.")
Wir wollen am Wochenende zur BII nach Hamburg
("We want to go to the BII in Hamburg during the weekend.")
Sie fahren nächste Woche zur LUU nach Hannover
("They are going to the LUU in Hannover next week.")

A short sample of spontaneous speech was elicited by asking subjects to describe a 500-m route through town from the speech lab, where the recordings were made, to a main university campus building. At the end of the study speakers received 10 euro for anonymous and voluntary participation and the purpose of the study was explained.

4. Acoustic analyses

The LPC formant tracking function in *Praat* (Boersma and Weenink, 2014) was used to estimate the first and second formant frequencies in the middle of the three last point vowel sequences of the abbreviations described above. Because of higher familiarity only the third recording of the sentence list was used. Formant tracking errors, especially occurring in F2 of /iː/ and /uː/, were manually corrected by the first author. Besides the single values for each vowel, mean F1 and F2 were estimated in line with Pierrehumbert *et al.* (2004) to index vowel space shift. All three F1 values per speaker (F1 of /aː/, /iː/, and /uː/) were added and divided by the number of vowels to get mean F1. The same procedure was applied to calculate mean F2. Furthermore, overall vowel space dispersion was calculated using the method proposed by Bradlow *et al.* (1996), equating the average Euclidian distances from the three tested vowels to the balance point of the vowel triangle. Finally, information for vowel durations was collected: measurement started at the midpoint of the intervowel glottalization phase and ended at the beginning of the following nasal sound. Median fundamental frequency was calculated from the sentence list and spontaneous speech samples of the sentence list using *Praat*'s pitch-tracking algorithm (autocorrelation method). We used median instead of mean fundamental frequency in order to better account for outliers. All acoustic measures were presented in the more perceptually relevant ERB values (Moore and Glasberg, 1983).

Given $\alpha = 0.05$ and 22 speakers per group, large between-groups main effects of $d = 0.80$ could be detected with a probability of $1 - \beta = 0.74$ (Cohen, 1977), which is nearly satisfactory. By implication, despite our relatively large sample, we were only able to detect large differences. In contrast to previous studies (Pierrehumbert *et al.*, 2004; Munson *et al.*, 2006; Rendall *et al.*, 2008; Camp, 2011; van Borsel *et al.*, 2013), we applied the conservative Bonferroni correction formula when multiple testing psychological and acoustic variables in case of replicating former findings (e.g., Do lesbians and straight women differ in a mean f0?). Hence, the overall probability of falsely claiming a correlation or group difference for each test is $\alpha = 0.05$. We did this in order to prevent claiming differences that are due to statistical shortcomings alone. Because this correction inflates the type II error, we used $\alpha = 0.05$ when we addressed a research question that has not been addressed before even in cases of multiple testing (e.g., What are the acoustic correlates of gender-role self-concept in women?). Because there is a cumulated probability of a type I error in these exploratory analyses, study 2 was carried out to reduce its probability.

B. Results

We first examined whether lesbian and straight speakers differed in any of the 12 psychological characteristics collected (see Table II: Jena columns) and adjusted the significance level to multiple tests ($\alpha = 0.0042$). Regarding present gender-role self-concept, lesbians and straight women differed in no measure. In terms of group affiliation, straight speakers were more strongly affiliated to the straight group than lesbians, whereas lesbians were more strongly affiliated to the lesbian group than straight women. Regarding present social environment, lesbians showed significantly higher current contact to other lesbians and reported a more equal distribution of lesbian and straight female friends than straight women did. Taken together, we found the expected differences in psychological characteristics related to sexual orientation except for gender-role self-concept.

1. Acoustic differences between lesbian and straight women

Unless stated otherwise, the following analyses were based on the categorization of speakers as either lesbian (Kinsey-like scores 1–3) or straight (Kinsey-like scores 5–7). Differences between lesbians and straight women were examined in 14 acoustic parameters (see Table III, adjusted significance level: $\alpha = 0.0036$). Whereas most of the acoustic differences were descriptively in the expected direction, with straight women demonstrating more gender conforming values than lesbians (except for F2 of /aː/, F2 of /uː/, and all three vowel durations), not a single significant difference was found (all $|ts[42]| \leq 2.40$, all $ps \geq 0.012$).

The failure to find significant acoustic differences could be due to our definition of the sexual orientation groups. Thus, we additionally tested for differences using previous studies' group definitions (Pierrehumbert *et al.*, 2004; Munson *et al.*, 2006; Rendall *et al.*, 2008; Camp, 2011) by comparing straight and non-straight women (i.e., lesbians and bisexual women). Those 20 women who identified themselves as mainly or exclusively straight on the Kinsey-like Scale formed the straight group (Kinsey-like scores 6–7), the remaining 31 women (Kinsey-like scores 1–5) were non-straight. Again, in none of the 14 investigated parameters did we find any significant differences (all $|ts[49]| \leq 1.52$, all $ps \geq 0.136$; see Table III). To conclude, no significant acoustic differences could be found between groups.

2. Differences within lesbians and within straight women

Contrasting two groups ignores any intragroup variability and gives a false impression of clear group demarcation. Thus, we next considered whether it is appropriate to treat lesbians (Kinsey-like scores 1–3) and straight women (Kinsey-like scores 5–7) as acoustically homogeneous entities. Because no previous study has investigated the acoustic correlates of sexual self-identification within lesbians and straight women (van Borsel *et al.*, 2013 had done so for lesbians only), the significance level was left unadjusted for exploratory reasons ($\alpha = 0.05$). As can be seen in Table III,

PROOF COPY [JASA-00586] 06706JAS
J_ID: JASMAN DOI: 10.1121/1.4988684 Date: 24-June-17 Stage: Page: 7 Total Pages: 17
ID: aipepub3b2server Time: 12:34 I Path: D:/AIP/Support/XML_Signal_Tmp/AI-JAS#170433

TABLE II. Mean differences between lesbians (Kinsey-like scores 1–3) and straight women (Kinsey-like scores 5–7) in psychological characteristics in Jena (study 1) and Berlin samples (study 2).

	Jena			Berlin			ANOVA[e]								
							Main effect sexual orientation			Main effect Sample			Interaction sexual orientation * Sample		
	$M_{lesbian}$[a]	$M_{straight}$[a]	p value[b]	$M_{lesbian}$[c]	$M_{straight}$[c]	p value[d]	F	p	η^2	F	p	η^2	F	p	η^2
Sexual orientation (objectifiable measures) towards...															
… men	**2.05**	**5.00**	**<0.001**	**2.12**	**5.54**	**<0.001**	**261.70**	**<0.001**	**0.74**	2.38	0.126	0.03	1.35	0.248	0.01
… women	**5.85**	**1.71**	**<0.001**	**6.11**	**2.22**	**<0.001**	**420.02**	**<0.001**	**0.82**	3.86	0.053	0.04	<1.00		
Gender-role self-concept															
TMF	4.48	5.12	0.039	4.57	5.25	0.014	**10.83**	**<0.001**	**0.10**	<1.00			<1.00		
GEPAQ-F[f]	4.27	4.25	0.833	4.05	4.16	0.459	<1.00			2.57	0.113	0.03	<1.00		
GEPAQ-M[f]	3.30	3.24	0.743												
CGNS[f]				2.99	2.54	0.116									
Group affiliation to...															
… lesbian group	**5.14**	**2.73**	**<0.001**	**4.70**	**2.41**	**<0.001**	**78.30**	**<0.001**	**0.45**	2.00	0.160	0.02	<1.00		
… straight group	**4.41**	**6.41**	**<0.001**	**3.89**	**6.04**	**<0.001**	**64.39**	**<0.001**	**0.41**	2.98	0.088	0.03	<1.00		
Psycho-social identification (CSES)[g]	4.52	4.94	0.147	4.55	4.60	0.876	1.11	0.296	0.01	<1.00			<1.00		
Social environment															
Current contact to lesbians	**5.18**	**3.18**	**<0.001**	**5.41**	**3.81**	**<0.001**	**54.42**	**<0.001**	**0.37**	3.11	0.081	0.03	<1.00		
Current contact to straight women	5.95	6.45	0.082	6.07	6.59	0.023	8.35	0.005	0.08	<1.00			<1.00		
Sexual orientation of friends	**4.55**	**6.33**	**<0.001**	**4.19**	**5.81**	**<0.001**	**80.96**	**<0.001**	**0.47**	5.35	0.023	0.05	<1.00		
Gender distribution of friends	4.95	4.64	0.405	4.85	4.56	0.242	1.95	0.165	0.02	<1.00			<1.00		
ChildCon to girls[h]				5.81	6.04	0.374									
ChildCon to boys				5.30	5.59	0.379									

[a] $n = 22$.
[b] The significance level is $\alpha = 0.0042$ after adjusting it to 12 tests according to Bonferroni correction formula. Significant p values are **bold**.
[c] $n = 27$.
[d] The significance level is $\alpha = 0.0036$ after adjusting it to 14 tests according to Bonferroni correction formula. Significant p values are **bold**.
[e] Because three lesbian-specific characteristics (coming-out extent, difficult process, and internalized homonegativity) were tested together with the eleven psychological characteristics measured for all Jena and Berlin women, significance level was adjusted to 14 tests: $\alpha = 0.0036$. Significant p values are **bold**.
[f] Scales for these psychological characteristics ranged from 1 to 5 (all others ranged from 1 to 7). German Extended Personality Attributes Questionnaire (GEPAQ). Endings indicate masculinity (-M) and femininity (-F) scales. Childhood Gender Nonconformity Scale (CGNS).
[g] Collective Self-Esteem Scale (CSES).
[h] Contact to corresponding gender during childhood (ChildCon).

TABLE III. Acoustic differences between lesbians and straight speakers (comparison 1) and between non-straight and straight speakers (comparison 2) as well as correlations of sexual orientation and acoustic parameters within lesbian and straight group in study 1.

	Comparison 1			Comparison 2			Correlations	
Acoustic parameter	Lesbian[a]	Straight[a]	*p* value[b]	Non-straight[c]	Straight[d]	*p* value[b]	$r_{Lesbian}$[a] (*p* value[e])	$r_{Straight}$[a] (*p* value[e])
Median f0 (ERB)								
spontaneous	5.40	5.48	0.528	5.42	5.50	0.439	**0.59 (0.004)**	0.24 (0.285)
read	5.45	5.56	0.395	5.42	5.59	0.136	**0.59 (0.004)**	0.28 (0.202)
Mean formant frequencies (ERB)								
F1	10.04	10.29	0.095	10.12	10.25	0.334	0.26 (0.238)	−0.14 (0.546)
F2	18.92	19.06	0.358	18.96	19.07	0.417	**0.53 (0.011)**	0.38 (0.080)
Single formant frequencies (ERB)								
F1								
/aː/	13.48	13.76	0.168	13.59	13.73	0.467	0.16 (0.477)	−0.03 (0.903)
/iː/	7.44	7.61	0.448	7.45	7.55	0.631	0.34 (0.116)	−0.19 (0.405)
/uː/	7.99	8.25	0.268	8.07	8.22	0.494	0.07 (0.765)	−0.10 (0.655)
F2								
/aː/	18.09	18.02	0.795	18.15	18.04	0.661	**0.51 (0.016)**	0.32 (0.150)
/iː/	23.09	23.41	0.069	23.18	23.42	0.157	0.30 (0.177)	0.23 (0.312)
/uː/	12.44	12.34	0.773	12.27	12.36	0.757	0.24 (0.279)	0.14 (0.536)
Vowel space dispersion (ERB)	13.24	13.67	0.051	13.39	13.67	0.176	0.09 (0.683)	0.12 (0.592)
Vowel duration (ms)								
/aː/	160.30	159.21	0.929	160.87	161.58	0.954	0.07 (0.729)	0.05 (0.833)
/iː/	164.78	132.63	0.020	152.12	135.82	0.252	−0.15 (0.511)	0.24 (0.282)
/uː/	176.35	158.19	0.164	170.69	158.03	0.321	−0.40 (0.066)	0.11 (0.613)

[a]$n = 22$ with Kinsey-like scores for lesbians 1–3 and for straight women 5–7.
[b]The significance level is $\alpha = 0.0036$ after adjusting it to 14 tests according to Bonferroni correction formula. Significant *p* values are **bold**.
[c]$n = 31$ with Kinsey-like scores for non-straight women 1–5.
[d]$n = 20$ with Kinsey-like scores for straight women 6–7.
[e]The significance level was left unadjusted due to exploratory reasons ($\alpha = 0.05$). Significant *p* values are **bold**.

no correlation of acoustic parameters and sexual orientation within the straight group was significant. However, the lesbian group showed four significant correlations: The more "exclusively lesbian" lesbians described themselves, the lower their median f0 in spontaneous ($r = 0.59$, $p = 0.004$; see Fig. 1) and read speech ($r = 0.59$, $p = 0.004$), mean F2 ($r = 0.53$, $p = 0.011$), and F2 for /aː/ ($r = 0.51$, $p = 0.016$). The correlation with mean F2 was caused by F2 of /aː/: Controlling for this variable the partial correlation was not significant ($r = 0.29$, $p = 0.21$).

3. Correlations of psychological characteristics and acoustic measures within the lesbian group

One of our main research questions is whether psychological characteristics influence lesbians' speech. Because the only previous study including psychological characteristics collected data of lesbians' coming-out age only (van Borsel *et al.*, 2013), significance level was left unadjusted for exploratory reasons. Some variables did not correlate with a single acoustic parameter: sexual orientation towards men, psycho-social identification, affiliation to lesbian group, current contact to straight women, internalized homonegativity, concealment motivation, and difficulty of the coming-out process. The psychological parameters depicted in Table IV were correlated significantly at least with one acoustic parameter.

Five psychological characteristics showed a significant correlation with only one acoustic parameter: sexual orientation towards women, affiliation to straight group, current contact to lesbians, sexual orientation of female friends, and coming-out extent. None of these five correlations is in line with gender inversion theory: the more gender conforming and heteronormative psychological characteristics lesbians

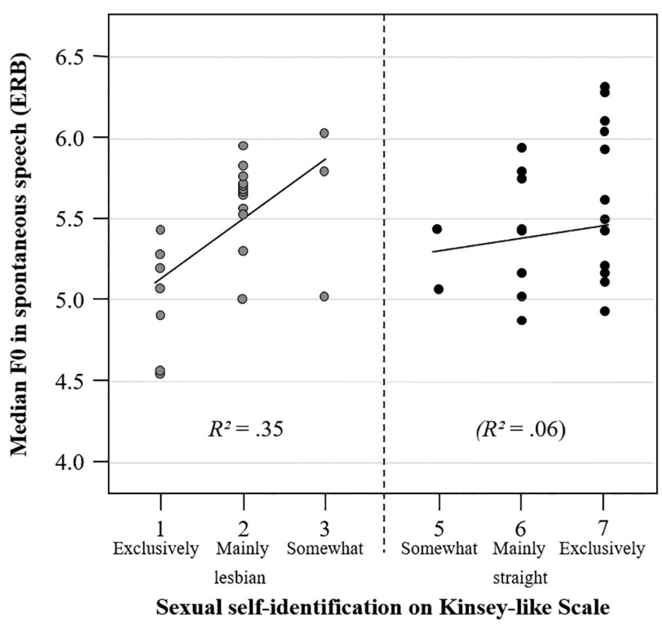


FIG. 1. Correlations of sexual self-identification and median f0 separately for straight women and lesbians in study 1. Non-significant determination coefficients are depicted in parentheses.

TABLE IV. Correlations of psychological characteristics and acoustic parameters within the lesbian group ($n = 22$) in study 1.[a]

Acoustic parameters	Sexual orientation towards women (p value)	Current contact to lesbians (p value)	Affiliation to straight group (p value)	Gender distribution of friends (p value)	Sexual orientation of female friends (p value)	Coming-out extent (p value)
Median f0 (ERB)						
spontaneous	−0.29 (0.193)	0.17 (0.446)	−0.04 (0.848)	0.24 (0.282)	−0.15 (0.499)	0.29 (0.206)
Read	−0.15 (0.499)	0.18 (0.437)	0.05 (0.831)	0.14 (0.546)	−0.16 (0.473)	0.24 (0.286)
Mean formant frequencies (ERB)						
F1	−0.07 (0.775)	0.12 (0.583)	−0.05 (0.837)	0.29 (0.196)	−0.11 (0.639)	0.37 (0.103)
F2	−0.35 (0.106)	−0.01 (0.951)	−0.31 (0.165)	−0.04 (0.859)	−0.23 (0.294)	0.22 (0.328)
Single formant frequencies (ERB)						
F1						
/aː/	0.18 (0.427)	**0.51 (0.014)**	−0.25 (0.271)	**0.57 (0.005)**	−0.26 (0.247)	**0.47 (0.030)**
/iː/	−0.20 (0.256)	−0.14 (0.542)	0.08 (0.719)	−0.01 (0.962)	−0.09 (0.679)	0.21 (0.366)
/uː/	−0.19 (0.394)	−0.33 (0.135)	0.18 (0.412)	−0.13 (0.554)	0.21 (0.518)	−0.03 (0.895)
F2						
/aː/	−0.16 (0.479)	0.22 (0.332)	**−0.43 (0.041)**	−0.04 (0.865)	−0.21 (0.349)	0.28 (0.217)
/iː/	−0.22 (0.326)	−0.08 (0.722)	−0.04 (0.856)	0.17 (0.464)	−0.35 (0.115)	0.13 (0.569)
/uː/	−0.40 (0.062)	−0.27 (0.227)	−0.17 (0.459)	**−0.44 (0.040)**	0.32 (0.154)	−0.03 (0.900)
Vowel space dispersion (ERB)	0.08 (0.736)	−0.14 (0.536)	0.05 (0.844)	**0.47 (0.028)**	**−0.46 (0.034)**	0.22 (0.349)
Vowel durations (ms)						
/aː/	−0.13 (0.566)	0.22 (0.320)	−0.24 (0.284)	0.25 (0.270)	−0.06 (0.784)	0.19 (0.415)
/iː/	**0.44 (0.043)**	0.29 (0.187)	−0.22 (0.333)	0.34 (0.124)	−0.09 (0.703)	−0.13 (0.564)
/uː/	−0.00 (0.997)	0.21 (0.353)	−0.24 (0.274)	0.33 (0.133)	−0.13 (0.557)	−0.12 (0.617)

[a]The significance level was left unadjusted due to exploratory reasons ($\alpha = 0.05$). Significant p values are **bold**.

reported (e.g., more affiliation to straight group), the less gender conforming were their speaking patterns (e.g., the lower their F2 of /aː/). Gender distribution of friends was the only psychological characteristic that was associated with more than one acoustic parameter ($n = 3$). Two of them corresponded to gender inversion theory, e.g., the more female friends lesbians reported, the more expanded their vowel spaces. Please note, however, that obtaining 8 significant results by chance when conducting 182 tests is not improbable, so all results should be treated with caution.

4. Acoustic correlates of present gender-role self-concept in the whole sample

To the best of our knowledge, there is no previous study dealing with the question whether there are any acoustic correlates of present gender-role self-concept. Hence, we left the significance level unadjusted. In our sample consisting of females only (thus holding biological effects of gender constant), we collected data on masculinity/femininity. Because the Traditional Masculinity-Femininity Scale (TMF) was shown to be better at differentiating lesbians and straight women compared to adjective-based measures of gender-role self-concept such as GEPAQ-M (Kachel *et al.*, 2016) which indicates its higher validity, we calculated correlations for TMF only.

Women reporting to be more feminine on TMF had higher mean F1 ($r = 0.37$, $p = 0.014$), mean F2 ($r = 0.39$, $p = 0.009$), F1 of /iː/ [$r = 0.31$, $p = 0.042$; see Fig. 2(a)] and F2 of /iː/ [$r = 0.38$, $p = 0.011$; see Fig. 2(b) and Table V]. When controlling for F1 and F2 of /iː/, the correlation of TMF with mean F1 and F2 was no longer significant ($r_{Mean\ F1} = 0.23$, $p = 0.137$; $r_{Mean\ F2} = 0.16$, $p = 0.300$) indicating that it was driven by F1 and F2 of /iː/. The correlation between TMF and F1 and F2 of /iː/ was driven by the lesbian sample, because no correlation coefficients became significant for the straight sample (F1 of /iː/: $r_{lesbian} = 0.43$, $p = 0.044$, $r_{straight} = 0.01$, $p = 0.979$; F2 of /iː/: $r_{lesbian} = 0.52$, $p = 0.012$, $r_{straight} = -0.13$, $p = 0.558$). The correlations for F1 and F2 of /iː/ were driven by one lesbian speaker [case number 11; see Figs. 2(a) and 2(b)] who produced outliers on TMF and F2 of /iː/ (>3 *SD*). After this speaker was excluded, no correlations were significant anymore.

C. Discussion

No significant acoustic differences between lesbians and straight women were found, even when the classification strategy of previous studies was used. However, we did find a significant acoustic correlate of sexual orientation (e.g., median f0 in spontaneous and read speech) *within* the lesbian but not within the straight group. Thus, we were able to demonstrate acoustically potent intragroup variance.

A small number of correlations between psychological characteristics and acoustic parameters were found for lesbians; however, these could be due to chance. For example, we cannot confidently deduce that having a higher proportion of female friends is accompanied by a more expanded vowel space within the lesbian group. Moreover, we cannot confidently deduce that there are acoustic correlates of present gender-role self-concept either for the overall sample or for the lesbian subsample, because correlations depended on the inclusion of one outlier.

To increase the speaker database and scrutinize the exploratory findings, we conducted a second study to get a clearer picture of the interplay between sexual orientation,

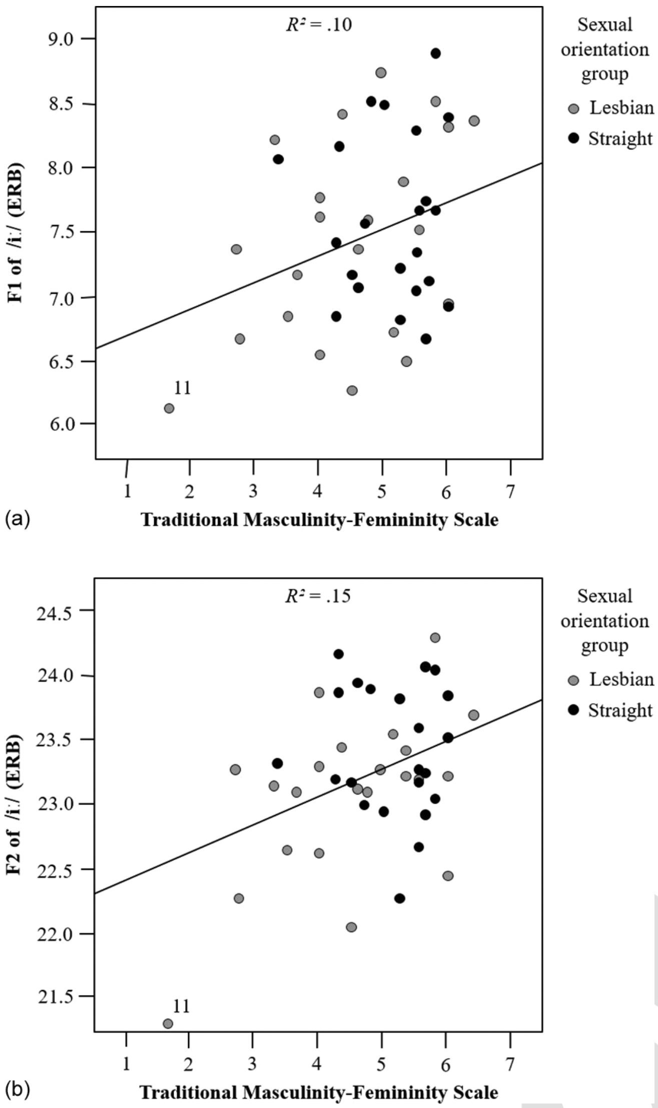


FIG. 2. (a) Correlation of present gender-role self-concept and F1 of /iː/ separately for lesbians and straight women in study 1. (b) Correlation of present gender-role self-concept and F2 of /iː/ separately for lesbians and straight women in study 1.

gender-role self-concept, and acoustic parameters. Can we confirm the tentative findings for intragroup variability? Is present gender-role self-concept indeed linked to acoustic measures?

III. STUDY 2

The replication study was carried out in Berlin. Additionally, the questionnaire was expanded to include gender-role self-concept during childhood, contact to other girls and boys, and perceived discrimination experiences.

A. Method

1. Participants

Recruitment strategies were adopted from study 1. Fifty-seven female native German speakers differing in sexual orientation were recruited. None reported any voice or speaking disorders. As in study 1, this sample was also highly homogeneous with regard to ethnicity, educational level, and age: nearly all women were Caucasian with a high educational level and ranged in age from 19 to 30 years ($M = 23.84$, $SD = 2.40$). The online questionnaire was again used to ensure a sample evenly distributed with regard to sexual orientation. Twenty-seven women described themselves as lesbian (Kinsey-like scores 1–3) and 27 women rated themselves as straight (scores 5–7). The three remaining women who identified themselves as equally lesbian and straight (score 4) were excluded from further analyses except for comparisons between non-straight and straight women (see Table I). No significant differences were found for any sociodemographic characteristics. Using a contingency coefficient test for two non-dichotomous nominally scaled variables (i.e., Cramérs phi test), almost the same proportion of straight ($n = 16$) and lesbian ($n = 13$) speakers were university students ($\phi_c = 0.23$, $p = 0.66$). There were also no differences between the Berlin and Jena women regarding any sociodemographic characteristics, except for graduation: straight women in Berlin were more likely to hold a university degree than straight women in Jena ($\phi_c = 0.47$, $p = 0.005$). Unsurprisingly, samples differed on dialect: most Jena women ($n = 29$) spent most of their lives in the south central German region (Thuringia, Saxony), whereas most Berlin women ($n = 24$) spent most of their lives in the north central German region (Saxony-Anhalt, Brandenburg, Berlin; $\phi_c = 0.55$, $p < 0.001$).

2. Psychological measures

All psychological characteristics were operationalized as in study 1 and showed similar internal consistencies (all Cronbach's αs ≥ 0.77) except for GEPAQ-M and concealment motivation which were excluded because of Cronbach's $\alpha < 0.7$. In addition to the psychological characteristics collected in study 1, information on gender-role self-concept during childhood, contact to boys and girls during childhood, and perceived discrimination experiences were elicited (see supplemental material for all items). Perceived discrimination experiences were reported by non-straight women only.

a. Gender-role self-concept during childhood. The Childhood Gender Nonconformity Scale (Rieger *et al.*, 2008) was used to assess how well a speaker considered herself to have rejected common gender stereotypes during childhood. It consisted of nine statements like "As a child I was called a tomboy by my parents and/or peers" and "As a child I preferred playing with boys rather than girls" that had to be rated on a 5-point scale ("1 = *strong disagreement*," "5 = *strong agreement*"; Cronbach's $\alpha = 0.88$). Higher scores indicated a higher degree of gender nonconformity during childhood.

b. Childhood social environment. Similar to current contact to lesbians and straight women, we assessed contact to girls and boys during childhood. We used two 7-point frequency-scales ("1 = *never*," "4 = *from time to time*," "7 = *always*"), one for contact with girls and one for contact with boys. Higher scores indicated more frequent contact to girls/boys during childhood.

TABLE V. Correlations of TMF Scale and acoustic measures for the overall sample (left side) and separated for the lesbian/straight subsamples (right side) in study 1 and 2.[a]

	Study 1			Study 2		
	Overall sample[b]	Subsamples[c]		Overall sample[b]	Subsamples[c]	
Acoustic parameter	Correlation with TMF (p value)	Lesbians (p value)	Straight women (p value)	Correlation with TMF (p value)	Lesbians (p value)	Straight women (p value)
Median f0 (ERB)						
spontaneous	0.18 (0.247)	0.19 (0.408)	0.12 (0.600)	**0.38 (0.005)**	**0.48 (0.011)**	0.21 (0.293)
read	0.20 (0.192)	0.17 (0.446)	0.17 (0.441)	**0.32 (0.018)**	0.36 (0.069)	0.16 (0.419)
Mean formant frequencies (ERB)						
F1	**0.37 (0.014)**	**0.43 (0.047)**	0.07 (0.749)	0.14 (0.310)	0.21 (0.297)	0.02 (0.909)
F2	**0.39 (0.009)**	**0.52 (0.013)**	0.07 (0.751)	0.20 (0.153)	0.10 (0.607)	0.21 (0.306)
Single formant frequencies (ERB)						
F1						
/aː/	0.21 (0.169)	0.24 (0.274)	−0.01 (0.981)	0.17 (0.208)	0.25 (0.216)	0.09 (0.660)
/iː/	**0.31 (0.042)**	**0.44 (0.043)**	−0.01 (0.996)	−0.08 (0.565)	−0.06 (0.775)	−0.13 (0.524)
/uː/	0.25 (0.104)	0.26 (0.252)	0.13 (0.563)	0.13 (0.369)	0.22 (0.267)	−0.02 (0.941)
F2						
/aː/	0.23 (0.127)	0.32 (0.148)	0.16 (0.481)	0.23 (0.099)	0.09 (0.640)	0.25 (0.203)
/iː/	**0.38 (0.011)**	**0.53 (0.011)**	−0.13 (0.558)	0.16 (0.245)	0.08 (0.696)	0.20 (0.320)
/uː/	0.05 (0.730)	0.02 (0.918)	0.15 (0.495)	−0.05 (0.716)	0.03 (0.876)	−0.20 (0.330)
Vowel space dispersion (ERB)	0.29 (0.053)	**0.45 (0.038)**	−0.23 (0.296)	0.21 (0.132)	0.12 (0.555)	0.24 (0.224)
Vowel duration (ms)						
/aː/	0.07 (0.633)	0.11 (0.653)	0.04 (0.874)	0.16 (0.249)	0.16 (0.432)	0.21 (0.310)
/iː/	0.02 (0.884)	0.26 (0.242)	−0.07 (0.746)	−0.13 (0.347)	−0.15 (0.475)	−0.02 (0.912)
/uː/	−0.15 (0.327)	0.09 (0.681)	−0.38 (0.080)	−0.03 (0.831)	−0.10 (0.618)	−0.05 (0.804)

[a]The significance level was left unadjusted due to exploratory reasons ($\alpha = 0.05$). Significant p values are **bold**. Traditional Masculinity-Femininity (TMF).
[b]$n = 44$.
[c]$n = 22$ each group.

c. Perceived discrimination experiences. We created a 13 item-encompassing scale in order to collect data on experiences of discrimination due to sexual orientation during the last five years. This measure is mainly based on two scales. Five of our scale's items were based on a translated and modified version of Herek and Berrill's (1990) Sample Survey of Anti-Gay Violence and Victimization (Steffens *et al.*, 2010). Four of them were decomposed halves of two items showing highest discrimination scores; e.g., the items "I was harassed" and "I feared something could happen to me" were based on the original one *Have you been harassed or did you fear something could happen to you?*. Six of our scale's items were based on the Everyday Discrimination Scale (Williams *et al.*, 1997) which was developed originally to assess soft forms of discrimination regarding race; e.g., the item "I was treated less politely compared to other people" based on the item *You are treated with less courtesy than other people are*. Moreover, we added two items: "I experienced disadvantages during renting a flat" and "Other people ignored me." All items of our scale were rated on a 5-point frequency-scale ("1 = *never*," "5 = *often*"; Cronbach's $\alpha = 0.79$). Higher scores indicated a higher degree of perceived discrimination.

3. Procedure and equipment

Again, data were collected in two phases. After the online questionnaire, sound recordings were conducted in a sound-proofed studio. In contrast to study 1, recordings were carried out by a female experimenter (Caucasian, mid-20-year-old, wearing a dark shirt and dark pants). Recordings were done with an AKG C1000S capacitor microphone connected to a preamplifying audio interface M-AUDIO Fast Track Ultra and a Windows Vista Lenovo T400 PC. The sampling rate was 44.1 kHz, the amplitude resolution 16 bit. Speakers recorded the same corpus of read speech, as well as a sample of spontaneous speech analogous to study 1 and were paid 18 euro. The same acoustic measurements were carried out as in study 1.

B. Results

We first examined whether any of the 14 psychological characteristics differed between lesbians (Kinsey-like scores 1–3) and straight women (scores 5–7); Bonferroni adjusted significance level: $\alpha = 0.0036$. All significant between-group differences from study 1 were significant again (i.e., affiliation to lesbian and straight group, current contact to lesbians, and sexual orientation of female friends; see Table II: Berlin columns). No measures of present gender-role self-concept and none of the newly added psychological measures were significant.

1. Acoustic differences between lesbians and straight women

Regarding differences between lesbians and straight women on acoustic measures, almost all values (except F2 in /uː/ and duration in /aː/ and /uː/) were descriptively higher

TABLE VI. Acoustic differences between lesbians and straight speakers (comparison 1) and between non-straight and straight speakers (comparison 2) as well as correlations of sexual orientation and acoustic parameters within lesbian and straight group in study 2.

	Comparison 1			Comparison 2			Correlations	
Acoustic parameter	Lesbian[a]	Straight[a]	*p* value[b]	Non-straight[c]	Straight[d]	*p* value[b]	$r_{Lesbian}$[a] (*p* value[e])	$r_{Straight}$[a] (*p* value[e])
Median f0 (ERB)								
spontaneous	5.21	5.32	0.341	5.20	5.37	0.138	−0.35 (0.075)	**0.43 (0.025)**
read	5.19	5.33	0.211	5.19	5.37	0.116	−0.32 (0.104)	0.31 (0.112)
Mean formant frequencies (ERB)								
F1	10.06	10.15	0.573	10.11	10.08	0.875	−0.14 (0.493)	−0.20 (0.315)
F2	18.67	18.84	0.153	18.75	18.77	0.870	−0.17 (0.391)	−0.24 (0.237)
Single formant frequencies (ERB)								
F1								
/aː/	13.76	13.91	0.504	13.82	13.89	0.759	−0.11 (0.597)	−0.10 (0.611)
/iː/	7.25	7.25	0.994	7.24	7.19	0.813	−0.05 (0.817)	−0.16 (0.419)
/uː/	7.77	7.79	0.889	7.84	7.63	0.309	−0.26 (0.195)	−0.23 (0.257)
F2								
/aː/	17.83	18.13	0.076	17.86	18.12	0.148	0.16 (0.433)	−0.01 (0.948)
/iː/	23.07	23.22	0.367	23.18	23.15	0.878	−0.20 (0.324)	−0.15 (0.470)
/uː/	11.49	11.47	0.969	11.58	11.27	0.304	−0.24 (0.232)	−0.29 (0.140)
Vowel space dispersion (ERB)	13.55	13.75	0.267	13.67	13.74	0.697	−0.15 (0.459)	−0.09 (0.668)
Vowel duration (ms)								
/aː/	160.05	158.70	0.892	157.95	159.69	0.859	−0.18 (0.401)	0.09 (0.678)
/iː/	154.62	144.96	0.423	153.60	167.85	0.299	−0.11 (0.593)	0.01 (0.964)
/uː/	154.05	165.30	0.377	155.74	141.06	0.338	0.18 (0.389)	0.32 (0.112)

[a]$n = 27$ with Kinsey-like scores for lesbians 1–3 and for straight women 5–7.
[b]The significance level is $\alpha = 0.0036$ after adjusting it to 14 tests according to Bonferroni correction formula. Significant *p* values are **bold**.
[c]$n = 36$ with Kinsey-like scores for non-straight women 1–5.
[d]$n = 21$ with Kinsey-like scores for straight women 6–7.
[e]The significance level was left unadjusted due to exploratory reasons ($\alpha = 0.05$). Significant *p* values are **bold**.

for straight women than for lesbians. However, with an adjusted significance level (14 tests: $\alpha = 0.0036$), none of the tested parameters differed significantly (see Table VI). Moreover, there was no acoustic difference when comparing non-straight (Kinsey-like scores 1–5) and straight women, as done in previous studies (Kinsey-like scores 6–7).

2. Differences within lesbians and within straight women

The significance level was left unadjusted for exploratory reasons ($\alpha = 0.05$). In contrast to study 1, significant correlations between sexual orientation and acoustic parameters could not be detected within the lesbian group, but were found within the straight group instead (see Table VI). As illustrated in Fig. 3, the straighter straight women called themselves, the higher their median f0 in spontaneous speech ($r = 0.43$, $p = 0.025$).

3. Correlations of psychological characteristics and acoustic measures within the lesbian group

Are there any significant correlations between psychological characteristics and acoustic parameters within the lesbian group? In addition to the characteristics examined in study 1, gender-role self-concept during childhood, contact to boys and girls during childhood, and perceived discrimination experiences were included. For exploratory reasons, the significance level was left at $\alpha = 0.05$. The psychological parameters depicted in Table VII were correlated significantly at least with one acoustic parameter.

Overall, 224 tests were conducted; 19 of them were significant, almost twice as many as would be expected by chance ($n = 11$). Similarly to study 1, sexual orientation

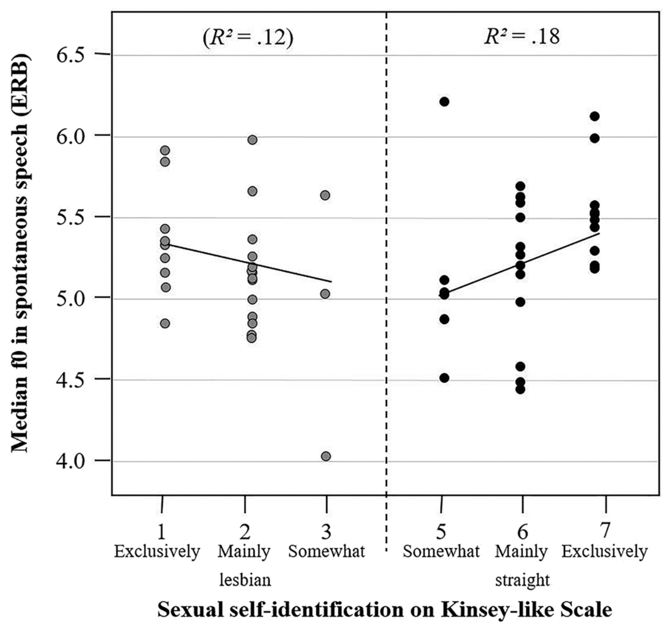


FIG. 3. Correlations of sexual self-identification and median f0 separately for straight women and lesbians in study 2. Non-significant determination coefficients are depicted in parentheses.

J_ID: JASMAN DOI: 10.1121/1.4988684 Date: 24-June-17 Stage: Page: 13 Total Pages: 17

TABLE VII. Correlations of psychological characteristics and acoustic parameters within the lesbian group ($n = 27$) in study 2.[a]

Acoustic parameters	Sexual orientation towards women (*p* value)	Psycho-social identification (CSES)[b] (*p* value)	Current contact to lesbians (*p* value)	Affiliation to lesbian group (*p* value)	Affiliation to straight group (*p* value)	Sexual orientation of female friends (*p* value)	Difficulty of coming-out process (*p* value)	Internalized homonegativity (*p* value)	CGNS[c] (*p* value)
Median f0 (ERB)									
spontaneous	0.14 (0.502)	−0.28 (0.165)	−0.07 (0.717)	0.11 (0.603)	0.16 (0.434)	−0.03 (0.884)	−0.21 (0.315)	−0.14 (0.474)	**−0.45 (0.019)**
read	0.18 (0.318)	−0.11 (0.570)	0.00 (0.997)	0.10 (0.616)	0.21 (0.306)	−0.15 (0.467)	−0.31 (0.129)	−0.22 (0.265)	**−0.48 (0.011)**
Mean formant frequencies (ERB)									
F1	0.23 (0.250)	0.13 (0.505)	−0.12 (0.568)	0.30 (0.127)	0.31 (0.118)	0.30 (0.128)	−0.32 (0.111)	**−0.45 (0.018)**	−0.31 (0.121)
F2	**−0.38 (0.048)**	−0.28 (0.158)	**−0.50 (0.008)**	0.04 (0.854)	0.07 (0.743)	**0.60 (0.001)**	−0.21 (0.293)	−0.02 (0.937)	−0.26 (0.196)
Single formant frequencies (ERB)									
F1									
/aː/	−0.06 (0.770)	0.07 (0.726)	−0.32 (0.107)	0.05 (0.824)	**0.44 (0.022)**	**0.39 (0.048)**	−0.13 (0.521)	**−0.45 (0.019)**	−0.15 (0.448)
/iː/	0.20 (0.307)	0.17 (0.405)	0.03 (0.887)	0.26 (0.199)	−0.07 (0.746)	0.19 (0.337)	−0.23 (0.259)	−0.22 (0.273)	−0.19 (0.340)
/uː/	**0.50 (0.007)**	0.05 (0.793)	0.17 (0.406)	**0.44 (0.021)**	0.20 (0.328)	−0.01 (0.948)	**−0.41 (0.040)**	−0.25 (0.206)	−0.36 (0.065)
F2									
/aː/	−0.28 (0.163)	**0.43 (0.024)**	−0.29 (0.144)	0.16 (0.421)	0.13 (0.513)	0.36 (0.064)	−0.35 (0.078)	**−0.50 (0.008)**	−0.18 (0.365)
/iː/	−0.30 (0.125)	**−0.44 (0.021)**	−0.30 (0.129)	0.04 (0.835)	0.02 (0.906)	**0.58 (0.002)**	−0.05 (0.828)	0.21 (0.306)	−0.22 (0.263)
/uː/	−0.16 (0.440)	−0.33 (0.089)	**−0.43 (0.026)**	−0.12 (0.567)	0.02 (0.941)	0.14 (0.482)	−0.15 (0.480)	0.07 (0.717)	−0.09 (0.674)
Vowel space dispersion (ERB)	−0.27 (0.170)	−0.37 (0.054)	−0.17 (0.396)	0.03 (0.870)	0.11 (0.572)	**0.55 (0.003)**	0.09 (0.652)	0.20 (0.311)	−0.15 (0.459)
Vowel durations (ms)									
/aː/	−0.13 (0.532)	−0.27 (0.199)	−0.18 (0.399)	−0.12 (0.580)	−0.12 (0.577)	0.11 (0.609)	0.23 (0.276)	0.01 (0.949)	−0.22 (0.282)
/iː/	0.03 (0.886)	−0.12 (0.550)	−0.34 (0.094)	−0.04 (0.851)	0.01 (0.945)	−0.34 (0.091)	0.08 (0.710)	0.05 (0.810)	0.03 (0.895)
/uː/	−0.12 (0.567)	0.04 (0.863)	0.01 (0.966)	**−0.41 (0.038)**	−0.27 (0.188)	−0.01 (0.964)	−0.17 (0.430)	0.02 (0.932)	0.20 (0340)

[a]The significance level was left unadjusted due to exploratory reasons ($\alpha = 0.05$). Significant *p* values are **bold**.
[b]Collective Self-Esteem Scale (CSES).
[c]Childhood Gender Nonconformity Scale (CGNS).

towards women, current contact to lesbians, affiliation to straight group, and sexual orientation of female friends were correlated with at least one acoustic parameter within lesbians; however, gender distribution of friends and coming-out extent were not significant. Of the newly included characteristics, only gender-role self-concept during childhood correlated with one acoustic parameter: the lower childhood gender nonconformity, the higher median f0 in spontaneous ($r = -0.45$, $p = 0.019$) and read speech ($r = -0.48$, $p = 0.011$).

4. Acoustic correlates of present gender-role self-concept in the whole sample

Considering lesbians (Kinsey-like scores 1–3) and straight women (scores 5–7) and leaving the significance level unadjusted for exploratory reasons, present gender-role self-concept showed acoustic correlates (see Table V). The more traditionally feminine women described themselves on TMF, the higher their median f0 in spontaneous ($r = 0.38$, $p = 0.005$; see Fig. 4) and read speech ($r = 0.32$, $p = 0.018$). However, no vowel-related correlations were found, not replicating study 1's findings concerning F1 and F2 of /iː/.

In study 1 correlations of present gender-role self-concept and acoustic parameters were due to the lesbian group. Hence, in study 2 we also computed all correlations for the straight and the lesbian group separately. In the straight group, no significant correlations with present gender-role self-concept and median f0 or any other acoustic measure could be observed (all $|rs| \leq 0.24$, all $ps \geq 0.24$). However, the more feminine lesbians called themselves, the higher their median f0 for spontaneous speech ($r = 0.48$, $p = 0.011$). As in study 1, correlations of TMF and median f0 in the overall group were driven by the lesbian group. But in contrast to study 1, the findings were not due to outliers.

5. Differences between Jena and Berlin women

Taken together, despite similarities in the size and composition of the samples, we obtained a number of differences in findings. In order to check if these differences were due to differences between samples in psychological characteristics and acoustic parameters, we conducted a series of 2 (sample:

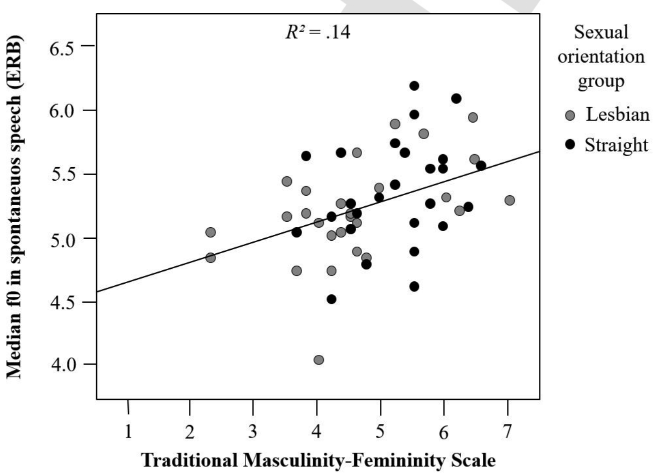


FIG. 4. Correlation of present gender-role self-concept and median f0 in spontaneous speech separately for lesbians and straight women in study 2.

Jena vs Berlin) $\times$ 2 (sexual orientation: lesbian vs straight) analysis of variances (ANOVAs) on the measures collected for all speakers. Additionally, we compared lesbians from Jena and Berlin regarding lesbian-specific psychological characteristics (adjusted significance levels: $\alpha = 0.0036$).

All psychological characteristics that differed between lesbians and straight women in the Jena and the Berlin sample were found to show significant effects of sexual orientation ($Fs[1,94] \geq 54.42$, $ps < 0.001$, $\eta^2 \geq 0.37$; see Table II): Affiliation to lesbian and straight group, GEPAQ-F, current contact to lesbians, and sexual orientation of female friends. Additionally, sexual orientation had a significant effect on TMF ($F[1, 94] = 10.83$, $p < 0.001$, $\eta^2 = 0.10$): Straight women called themselves more feminine/less masculine than lesbians. Psycho-social identification, current contact to straight women, and gender distribution of friends showed no significant sexual orientation effect ($Fs[1, 94] \leq 8.35$, $ps \geq 0.005$, $\eta^2 \leq 0.08$). There was no significant effect of sample (all $Fs[1, 93] \leq 5.35$, all $ps \geq 0.023$, all $\eta^2 \leq 0.05$) or interaction of the two factors (all $Fs[1, 94] \leq 1.35$, all $ps \geq 0.248$, all $\eta^2 \leq 0.01$). Regarding lesbian-specific psychological characteristics, no significant difference between lesbians from Jena and Berlin could be observed (all $|ts[46]| \leq 2.70$, all $ps \geq 0.010$).

In the ANOVAs for acoustic characteristics, we found a main effect of sample for one acoustic parameter (all other $Fs[1, 94] \leq 6.33$, all other $ps \geq 0.014$, all other $\eta^2 \leq 0.06$): Jena women scored higher than Berlin women on F2 in /uː/ ($M_{Jena} = 12.39$, $M_{Berlin} = 11.48$; $F[1, 94] = 16.09$, $p < 0.001$, $\eta^2 = 0.15$). No main effect of sexual orientation (all $Fs[1, 94] \leq 5.25$, all $ps \geq 0.024$, all $\eta^2 \leq 0.05$) or interaction effect was significant (all $Fs[1, 94] \leq 1.50$, all $ps \geq 0.224$, all $\eta^2 \leq 0.02$).

C. Discussion

Replicating study 1, no acoustic differences were found between lesbians and straight women, even if the classification strategy from previous studies was applied. However, in line with study 1, we found evidence that lesbians and straight women cannot be considered internally homogenous groups. Although in contrast to study 1, no correlations of sexual orientation and acoustic parameters were found for the lesbian group, there was a correlation with median f0 in spontaneous speech for the straight group.

Moreover, intragroup variability for lesbians regarding various psychological characteristics is consistent with study 1. Sexual orientation towards women, current contact to lesbians, affiliation to straight group, and sexual orientation of female friends showed acoustic correlates again. Regarding the psychological data additionally collected in study 2, only childhood gender nonconformity showed acoustic correlates: The more gender conforming lesbians described themselves during childhood, the higher the median f0. Hence, variance within the lesbian group was demonstrated again.

As in study 1, acoustic correlates of gender-role self-concept were found. However, this time speakers did not use vowel space characteristics but median f0 independent of speech mode (spontaneous vs read speech) to reflect their masculinity/femininity. In line with study 1 it was driven by lesbians. However, it was not due to outliers. Hence, present gender-role self-concept was shown to have acoustic correlates for lesbians.

One main difference occurred in result patterns: We found an opposite effect of sexual orientation on the intragroup level, with straight women but not lesbians differing in acoustic parameters. Different result patterns between the two samples can only partially be attributed to acoustic and psychological differences between the Jena and Berlin women. Jena women produced higher F2 values for /uː/ than the Berlin women did. However, no between-sample differences in psychological characteristics occurred.

IV. CONCLUSIONS

The present study represents a large-scale investigation of the acoustic correlates of women's sexual orientation, using a German-speaking sample, applying fine-grained measurement of sexual orientation, and including psychological characteristics. As shown in two studies, no between-group differences were found for German lesbians and straight women in speaking patterns, even when the classification strategy of previous studies was used (non-straight vs straight). That challenges gender inversion theory (Kite and Deaux, 1987). Our findings on median f0 are not in line with studies focusing on connected speech that converged on the finding that straight women showed higher mean f0 than lesbians (Camp, 2011; van Borsel *et al.*, 2013), but replicated null findings of studies that investigated vowel-specific mean f0 values (Rendall *et al.*, 2008; Munson *et al.*, 2006). Moreover, we found no significant differences in vowel space characteristics, in line with Munson *et al.* (2006). However, our results are not consistent with studies that demonstrated higher mean F1 and F2 values for straight compared to non-straight women (Pierrehumbert *et al.*, 2004; Rendall *et al.*, 2008). Sexual orientation classification strategy had no impact on result patterns.

Although there were no between-group differences, we found several pieces of evidence for acoustically potent within-group differences. First, a more detailed analysis of sexual orientation yielded evidence for within-group acoustic variability: The more exclusively lesbian lesbians described themselves, the lower their F2 in /aː/ as well as median f0 in read and spontaneous speech (study 1: male experimenter), which contradicts van Borsel *et al.* (2013). Similarly, the more exclusively straight straight women called themselves, the higher their median f0 in spontaneous speech (study 2: female experimenter). Summing up, we recommend the use of fine-grained sexual assessment scales, such as continuous, one-dimensional, bipolar measures (e.g., Kinsey-Scale) and the use of continuous two dimensional, unipolar measures separating sexual orientation towards men and women and thus also allowing for a phonetic investigation of bisexual and asexual women. Moreover, we recommend a speech accommodation experiment manipulating experimenter gender and investigating its impact on speakers' median f0, depending on speakers' sexual orientation, for future research.

Further evidence for acoustically potent within-group differences was found when focusing on lesbians. The second piece of evidence was demonstrated when looking at actual masculinity/femininity. Hitherto, acoustic studies on gendered speech have often concentrated on biological differences, directing less attention to social components of gender (see Smyth and Rogers, 2008). To our knowledge, ours is the first study investigating whether social components of gender are reflected in speech as well. We found acoustic correlates of lesbians' present gender-role self-concept. The more feminine/less masculine lesbians described themselves, the higher their median f0 in spontaneous and read speech (study 2). This finding supports gender inversion theory given that a higher degree of gender conformity is accompanied by more gender conforming speaking patterns.

The third piece of evidence for acoustically potent within-group differences was obtained when looking at other psychological characteristics of lesbians. Sexual orientation towards women, current contact to lesbians, affiliation to straight group, and sexual orientation of female friends have repeatedly shown acoustic correlates. Some findings support stereotypical association of straightness and gender conforming speaking patterns (study 1: the higher lesbians' sexual orientation towards women, the lower their mean F2; study 2: the more female friends lesbians report, the higher their F1 in /aː/), while others contradict it (study 1: the more affiliation lesbians reported to straight women, the lower their F2 in /aː/; study 2: the more female friends lesbians report, the lower their F2 in /iː/).

All three pieces of acoustically potent within-group variability indicate that straight women and especially lesbians cannot be considered as internally homogeneous groups. Overall, this finding contradicts gender inversion theory because lesbians (and straight women) were stereotypically assumed to form a homogeneous group with respect to psychological characteristics and speaking patterns. Future research should investigate the intragroup variability of a male sample and include psychological characteristics, especially present gender-role self-concept, as well.

When dealing with acoustic correlates of sexual orientation in women, the question arises what constitutes lesbian/straight speech, and we need to decide between two perspectives: "linguistics of community" vs "linguistics of contact" (Barrett, 1997). The first perspective focuses on language/speech communities which seem to coincide with social categories (Barrett, 1997). This is what we did implicitly when conducting our research on lesbians and straight women. The second perspective counters that language communities are not externally definable categories because it is unclear which characteristics someone ought to possess to be seen as a social category member (see also Queen, 1997, on who is a true lesbian; Zwicky, 1997). It highlights that social categories are not homogeneous and that they do not strongly differ from each other (Levon, 2006; Zwicky, 1997), and acknowledges that there are intraindividual differences in speech patterns, for instance, depending on the communicative context (i.e., topic of the conversation, location, listeners etc.; Waksler, 2001; Levon, 2006). We used sexual self-categorization in order to circumvent externally attributed community memberships (see Smyth *et al.*, 2003), but provided evidence that sexual self-categorization and more objectifiable measures of sexual orientation seem to represent the same underlying construct. We sub-classified lesbians and straight women using a fine-grained sexual orientation measurement and other psychological characteristics to account for intragroup variability. Hence, our approach integrated both perspectives.

Although we used a relatively large sample size, test power in our analysis was low and was lowered further by adjusting the significance level to the number of tests, except in exploratory analyses. However, a high risk of not detecting actual acoustic differences between two groups is more justifiable when dealing with a highly sensitive topic as sexual orientation. It is important to know the probability of falsely claiming differences between lesbians and straight women and thereby unnecessarily contributing to a scientific foundation of speech and voice stereotypes (see Sec. I). Hence, we recommend that future studies rely on even larger numbers of lesbians and straight women ($n = 100$ per group) and simultaneously lowering the number of investigated psychological and acoustic characteristics to assure high test power.

As in every study dealing with sexual orientation, the generalization of the results is restricted. We do not know how acoustic, psychological, or other characteristics are distributed within sexual minority populations (Sandfort, 1997). Additionally, it is very likely that sexual minority members who are comfortable with their sexual orientation are more likely to participate in a study explicitly calling for lesbians and bisexual women. Therefore, self-selection and priming effects induced by a direct reference to sexual orientation are major problems for all studies on sexual minorities.

Our recordings were done in a laboratory by experimenters with prior knowledge of speakers' sexual orientation and analyzed with regard to acoustic parameters commonly used in studies on the speech of sexually divergent women. We are well aware that (1) experimenters may have subconsciously treated speakers differently depending on their sexual orientation, (2) recording spontaneous speech in more natural surroundings may elicit more representative or extreme patterns (Podesva, 2011; see also Waksler, 2011), and (3) two studies focused on other acoustic parameters such as spectral energy distribution in /s/ and /ʃ/ (Munson *et al.*, 2006) and speaking rate (Munson, 2010). Future studies should rely on experimenters blind to speakers' sexual orientation, reduce artifacts of laboratory speech by investigating natural surroundings by using mobile microphones (Podesva, 2011), and investigate other acoustic correlates worthy of scrutiny, such sibilant features, speaking rate, and voice quality (e.g., breathiness; Simpson, 2009; Henton and Bladon, 1985).

Undoubtedly, the most important general finding of the present study, which will inform future studies, is that differences between the group categorizations that have often been adopted in previous studies (e.g., lesbian vs straight, straight vs non-straight) are less important than differences found *within* these groups. Given that, it becomes less easy to answer the question how a lesbian (or a straight woman) speaks because intragroup variability erodes stereotypes.

ACKNOWLEDGMENTS

The current research was funded by a grant from the German Research Foundation to the third and second authors (DFG STE 938/10-2; FOR 1097). We thank Jörg Dreyer, Dr. Susanne Fuchs, and Dr. Stefanie Jannedy from the Center for Applied General Linguistics in Berlin for providing a phonetic laboratory, we thank Karoline Nestler for help with data collection and Felicia Schuld and Felix Sokoll for data pre-processing. Moreover, we kindly thank JASA's responsible associate editor Laura L. Koenig and anonymous reviewers for valuable comments on previous versions of this manuscript.

[1]See supplementary material at http://dx.doi.org/10.1121/1.4988684 for detailed overview about psychological measures.

Baeck, H., Corthals, P., and van Borsel, J. (**2011**). "Pitch characteristics of homosexual males," J. Voice **25**, 211–214.

Bailey, J. M., and Zucker, K. J. (**1995**). "Childhood sex-typed behavior and sexual orientation: A conceptual analysis and quantitative review," Dev. Psychology **31**, 43–55.

Barrett, R. (**1997**). "The 'homo-genius' speech community," in *Queerly Phrased. Language, Gender, and Sexuality*, edited by A. Livia and K. Hall (Oxford University Press, New York), pp. 181–201.

Boersma, P., and Weenink, D. (**2014**). "PRAAT: Doing phonetics by computer (version 5.3.71) [computer code]," http://www.praat.org/ (Last viewed September 18, 2014).

Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (**1996**). "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics," Speech Comm. **20**, 255–272.

Camp, M. (**2011**). "Japanese lesbian speech: Sexuality, gender identity, and language," Ph.D. thesis, University of Arizona, Columbus, OH, 216 pp.

Cohen, J. (**1977**). *Statistical Power Analysis for the Behavioral Sciences* (Lawrence Erlbaum Associates, Mahwah, NJ), 567 pp.

Eckert, P. (**1990**). "The whole woman: Sex and gender differences in variation," Lang. Variation Change **1**, 245–267.

Henton, C. G., and Bladon, R. A. W. (**1985**). "Breathiness in normal female speech: Inefficiency versus desirability," Lang. Commun. **5**, 221–227.

Herek, G. M., and Berrill, K. T. (**1990**). "Documenting the victimization of lesbians and gay men: Methodological issues," J. Interpersonal Violence **5**, 301–315.

Herek, G. M., Gillis, J. R., Cogan, J. C., and Glunt, E. K. (**1997**). "Hate crime victimization among lesbian, gay, and bisexual adults: Prevalence, psychological correlates, and methodological issues," J. Interpersonal Violence **12**, 195–215.

Hodson, G., Harry, H., and Mitchell, A. (**2009**). "Independent benefits of contact and friendship on attitudes toward homosexuals among authoritarians and highly identified heterosexuals," Eur. J. Soc. Psychol. **39**, 509–525.

Jacobs, G. (**1996**). "Lesbian and gay male language use: A critical review of the literature," Am. Speech **71**, 49–71.

Jeffreys, S. (**1999**). "Bisexual politics: A superior form of feminism?," Women's Studies Int. Forum **22**, 273–285.

Kachel, S., Steffens, M. C., and Niedlich, C. (**2016**). "Traditional masculinity and femininity: Validation of a new scale assessing gender roles," Front. Psychol. **7**, 1–19.

Kinsey, A., Pomeroy, W., Martin, C., and Gebhard, P. (**1953**). *Sexual Behavior in the Human Female* (Saunders, Philadelphia, PA), 842 pp.

Kite, M., and Deaux, K. (**1987**). "Gender belief systems: Homosexuality and the implicit inversion theory," Psychol. Women Quart. **11**, 83–96.

Klein, F., Sepekoff, B., and Wolf, T. I. (**1985**). "Sexual orientation: A multi-variable dynamic process," J. Homosexuality **11**, 35–49.

Kohler, K. J. (**1994**). "Glottal stops and glottalization in German. Data and theory of connected speech processes," Phonetica **51**, 38–51.

Kulick, D. (**2000**). "Gay and lesbian language," Annu. Rev. Anthropol. **29**, 243–285.

Levon, E. (**2006**). "Hearing 'gay': Prosody, interpretation, and the affective judgments of men's speech," Am. Speech **81**, 56–78.

Lippa, R. A. (**2005**). "Sexual orientation and personality," Annu. Rev. Sex Res. **16**, 119–153.

Loulan, J. (**1990**). *The Lesbian Erotic Dance: Butch, Femme, Androgyny, and Other Rhythms* (Spinsters Ink., Minneapolis, MN), 287 pp.

Luhtanen, R., and Crocker, J. (**1992**). "A collective self-esteem scale: Self-evaluation of one's social identity," Pers. Soc. Psychol. Bull. **18**, 302–318.

Mohr, J. J., and Kendra, M. S. (**2011**). "Revision and extension of a multidimensional measure of sexual minority identity: The Lesbian, Gay, and Bisexual Identity Scale," J. Counseling Psychol. **58**, 234–245.

Moonwomon-Baird, B. (**1997**). "Toward a study of lesbian speech," in *Queerly Phrased. Language, Gender, and Sexuality*, edited by A. Livia and K. Hall (Oxford University Press, New York), pp. 202–213.

Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.

Munson, B. (**2010**). "The influence of actual and perceived sexual orientation on diadochkinetic rate in women and men," in *Proceedings of the 11th Annual Conference of the International Speech Communication Association*, pp. 1525–1528.

Munson, B. (**2011**). "Lavender lessons learned: Or, what sexuality can teach us about phonetic variation," Am. Speech **86**, 14–31.

Munson, B., and Babel, M. (**2007**). "Loose lips and silver tongues, or, projecting sexual orientation through speech," Lang. Linguist. Compass **1**, 416–449.

Munson, B., McDonald, E. C., DeBoe, N. L., and White, A. R. (**2006**). "The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech," J. Phonetics **34**, 202–240.

Pearcey, S. M., Docherty, K. J., and Dabbs, J. M., Jr. (**1996**). "Testosterone and sex role identification in lesbian couples," Physiol. Behav. **60**, 1033–1035.

Pierrehumbert, J. B., Bent, T., Munson, B., Bradlow, A. R., and Bailey, M. J. (**2004**). "The influence of sexual orientation on vowel production (L)," J. Acoust. Soc. Am. **116**, 1905–1908.

Podesva, R. J. (**2011**). "The California vowel shift and gay identity," Am. Speech **86**, 32–51.

Queen, R. M. (**1997**). " 'I don't speak spritch.' Locating lesbian language," in *Queerly Phrased. Language, Gender, and Sexuality*, edited by A. Livia and K. Hall (Oxford University Press, New York), pp. 233–256.

Rendall, D., Vasey, P. L., and McKenzie, J. (**2008**). "The queen's English: An alternative, biosocial hypothesis for the distinctive features of 'gay speech,' " Arch. Sexual Behav. **37**, 188–204.

Rieger, G., Linsenmeier, J. A. W., Gygax, L., and Bailey, J. M. (**2008**). "Sexual orientation and childhood gender nonconformity: Evidence from home videos," Dev. Psychol. **44**, 46–58.

Runge, T. E., Frey, D., Gollwitzer, P. E., Helmreich, R. L., and Spence, J. T. (**1981**). "Masculine (instrumental) and feminine (expressive) traits: A comparison between students in the United States and West Germany," J. Cross-Cultural Psychol. **12**, 142–162.

Sandfort, T. G. M. (**1997**). "Sampling male homosexuality," in *Researching Sexual Behavior: Methodological Issues*, edited by J. Bancroft (Indiana University Press, Bloomington, IN), pp. 261–275.

Schubert, T. W., and Otten, S. (**2002**). "Overlap of self, ingroup, and outgroup: Pictorial measures of self-categorization," Self Ident. **1**, 353–376.

Shively, M. G., Jones, C., and DeCecco, J. P. (**1985**). "Research on sexual orientation: Definitions and methods," J. Homosexuality **9**, 127–136.

Simpson, A. P. (**2009**). "Phonetic differences between male and female speech," Lang. Linguist. Compass **3**, 621–640.

Singh, D., Vidaurri, M., Zambarano, R. J., and Dabbs, J. M., Jr. (**1999**). "Lesbian erotic role identification: Behavioral, morphological, and hormonal correlates," J. Personality Soc. Psychol. **76**, 1035–1049.

Smyth, R., Jacobs, G., and Rogers, H. (**2003**). "Male voices and perceived sexual orientation: An experimental and theoretical approach," Lang. Soc. **32**, 329–350.

Smyth, R., and Rogers, H. (**2008**). "Do gay-sounding men speak like women?," Toronto Working Pap. Linguist. **27**, 129–144.

Spence, J. T., and Helmreich, R. L. (**1978**). *Masculinity and Femininity: Their Psychological Dimensions, Correlates and Antecedents* (University of Texas Press, Austin, TX), 297 pp.

Steffens, M. C., Bergert, M., and Heinecke, S. (**2010**). *Doppelt diskriminiert oder gut integriert? Zur Lebenssituation von Lesben and Schwulen mit Migrationshintergrund* (*Discriminated twice or well integrated? The living situation of lesbians and gay men with a migration background*) (LSVD, Berlin, Germany), 51 pp.

Steffens, M. C., and Wagner, C. (**2004**). "Attitudes toward lesbians, gay men, bisexual women, and bisexual men in Germany," J. Sex Res. **41**, 137–149.

Stuart-Smith, J. (**2007**). "Empirical evidence for gendered speech production: /s/ in Glaswegian," in *Laboratory Phonology*, edited by J. Cole and J. I. Hualde (Mouton de Gruyter, New York), pp. 65–86.

Szymanski, D. M., and Chung, B. Y. (**2001**). "The internalized homophobia scale for lesbians: A rational/theoretical approach," J. Homosexuality **41**, 37–51.

van Borsel, J., de Bruyn, E., Lefebvre, E., Sokoloff, A., de Ley, S., and Baudonck, N. (**2009**). "The prevalence of lisping in gay men," J. Comm. Disord. **42**, 100–106.

van Borsel, J., Vandaele, J., and Corthals, P. (**2013**). "Pitch and pitch variation in lesbian women," J. Voice **27**, 656.e13–656.e16.

Waksler, R. (**2001**). "Pitch range and women's sexual orientation," Word **52**, 69–77.

Walters, S. D. (**1996**). "From here to queer: Radical feminism, postmodernism, and the lesbian menace (or why can't a woman be more like a fag?)," Signs **21**, 830–869.

Weirich, M., and Simpson, A. P. (**2013**). "Investigating the relationship between average speaker fundamental frequency and acoustic vowel space size," J. Acoust. Soc. Am. **134**, 2965–2974.

Williams, D. R., Yu, Y., Jackson, J. S., and Anderson, N. B. (**1997**). "Racial differences in physical and mental health: Socioeconomic status, stress, and discrimination," J. Health Psychol. **2**, 335–351.

Worthington, R. L., and Reynolds, A. L. (**2009**). "Within-group differences in sexual orientation and identity," J. Counseling Psychol. **56**, 44–55.

Zwicky, A. M. (**1997**). "Two lavender issues for linguists," in *Queerly Phrased. Language, gender, and sexuality*, edited by A. Livia and K. Hall (Oxford University Press, New York), pp. 21–34.

Manuscript 2 – Related Appendix

ONLINE APPENDIX: SCALE DESCRIPTIONS

Sexual orientation

Sexual self-identification (Kinsey-like Scale).*

Item: “Regarding my sexual orientation I identify as...”.

1	2	3	4	5	6	7
Exclusively lesbian	Mainly lesbian	Somewhat lesbian	Equally lesbian and straight	Somewhat straight	Mainly straight	Exclusively straight

*The scale used in the present study is based on the Kinsey-Scale (Kinsey, Pomeroy, Martin, and Gebhard, 1953) that has been adapted in several ways. We eliminated language bias (“lesbian” instead of “homosexual”), used more comprehensible scale descriptions (“mainly straight” instead of “predominantly heterosexual, only incidentally homosexual”) by keeping the equidistance of the scale points, used scores ranging from 1 to 7 instead of 0 to 6, and re-poled the scale in order to have a higher straightness indicated by higher scores.

S*exual orientation towards men and women (more objectifiable measure).*

“How often have you experienced…?”

	Never	Very seldom	Seldom	From time to time	Often	Very often	Always
	1	*2*	*3*	*4*	*5*	*6*	*7*
…sexual fantasies with men	○	○	○	○	○	○	○
…sexual fantasies with women	○	○	○	○	○	○	○
…romantic attraction towards men	○	○	○	○	○	○	○
…romantic attraction towards women	○	○	○	○	○	○	○
…physical attraction towards men	○	○	○	○	○	○	○
…physical attraction towards women	○	○	○	○	○	○	○
…sexual interaction with men	○	○	○	○	○	○	○
…sexual interaction with women	○	○	○	○	○	○	○

Present gender-role self-concept

Traditional Masculinity-Femininity Scale (TMF by Kachel, Steffens, and Niedlich, 2016).

	Very masculine						Very feminine
	1	*2*	*3*	*4*	*5*	*6*	*7*
I consider myself…	○	○	○	○	○	○	○
Ideally, I would like to be…	○	○	○	○	○	○	○
Traditionally, my interests would be regarded as…	○	○	○	○	○	○	○
Traditionally, my attitudes and beliefs would be regarded as…	○	○	○	○	○	○	○
Traditionally, my behavior would be regarded as…	○	○	○	○	○	○	○
Traditionally, my outer appearance would be regarded as…	○	○	○	○	○	○	○

German Extended Personal Attributes Questionnaire Masculinity-Scale (GEPAQ-M by Runge, Frey, Gollwitzer, Helmreich, and Spence, 1981).

Not independent	*1*	*2*	*3*	*4*	*5*	Very independent
Very passive	*1*	*2*	*3*	*4*	*5*	Very active
Not competitive	*1*	*2*	*3*	*4*	*5*	Very competitive
*Decisive	*1*	*2*	*3*	*4*	*5*	Not decisive
Gives up easily	*1*	*2*	*3*	*4*	*5*	Never gives up
Not self-confident	*1*	*2*	*3*	*4*	*5*	Self-confident
Feels inferior	*1*	*2*	*3*	*4*	*5*	Feels superior
Doesn’t stand up under pressure	*1*	*2*	*3*	*4*	*5*	Stands up under pressure

*Items with asterisks had to be recoded (1 to 5, 2 to 4, 3 to 3, 4 to 2, 5 to 1).

German Extended Personal Attributes Questionnaire Femininity-Scale (GEPAQ-F by Runge, Frey, Gollwitzer, Helmreich, and Spence, 1981).

Not emotional	*1*	*2*	*3*	*4*	*5*	Very emotional
*Devotes self to others	*1*	*2*	*3*	*4*	*5*	Doesn't devote self to others
Very rough	*1*	*2*	*3*	*4*	*5*	Very gentle
Not helpful	*1*	*2*	*3*	*4*	*5*	Very helpful
Very unkind	*1*	*2*	*3*	*4*	*5*	Very kind
Not aware of feelings	*1*	*2*	*3*	*4*	*5*	Aware of feelings
Not understanding	*1*	*2*	*3*	*4*	*5*	Very understanding
Cold	*1*	*2*	*3*	*4*	*5*	Warm

*Items with asterisks had to be recoded (1 to 5, 2 to 4, 3 to 3, 4 to 2, 5 to 1).

Gender-role self-concept during childhood (Study 2 only)

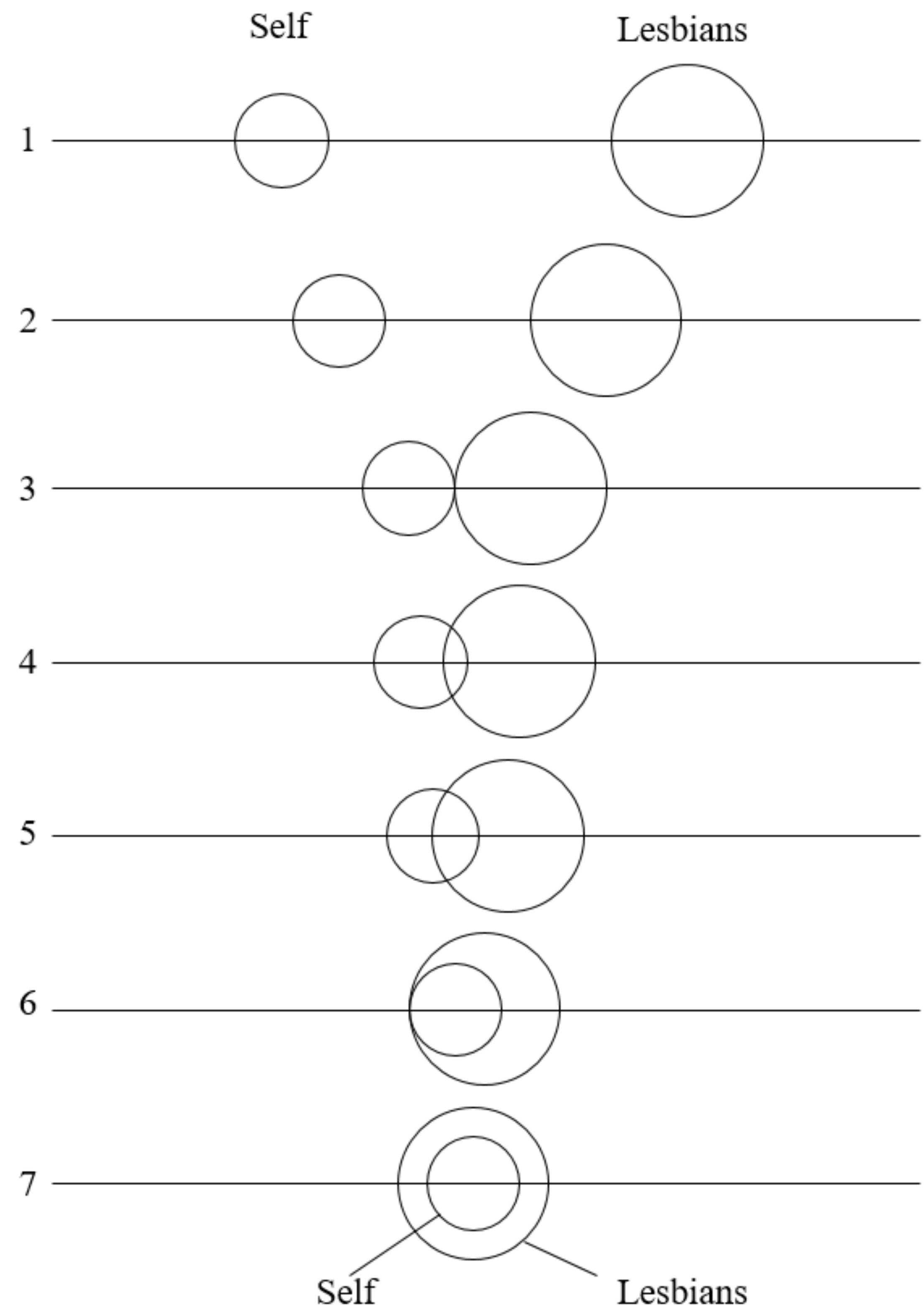
Childhood Gender Nonconformity Scale (CGNS by Rieger, Linsenmeier, Gygax, and Bailey, 2008).

	I strongly disagree		Partly partly		I strongly agree
	1	*2*	*3*	*4*	*5*
I was a masculine girl.	○	○	○	○	○
As a child I was called a tomboy by my parents and/or peers.	○	○	○	○	○
As a child I sometimes wished I had been born a boy rather than a girl.	○	○	○	○	○
As a child I preferred playing with boys rather than girls.	○	○	○	○	○
As a child I play typical boys' games.	○	○	○	○	○
As a child I often felt that I had more in common with boys than girls.	○	○	○	○	○
As a child I usually avoided feminine clothing.	○	○	○	○	○
As a child I liked competitive sports such as soccer or basketball.	○	○	○	○	○
* As a child I behaved like a typical girl.	○	○	○	○	○

*Items with asterisks had to be recoded (1 to 5, 2 to 4, 3 to 3, 4 to 2, 5 to 1).

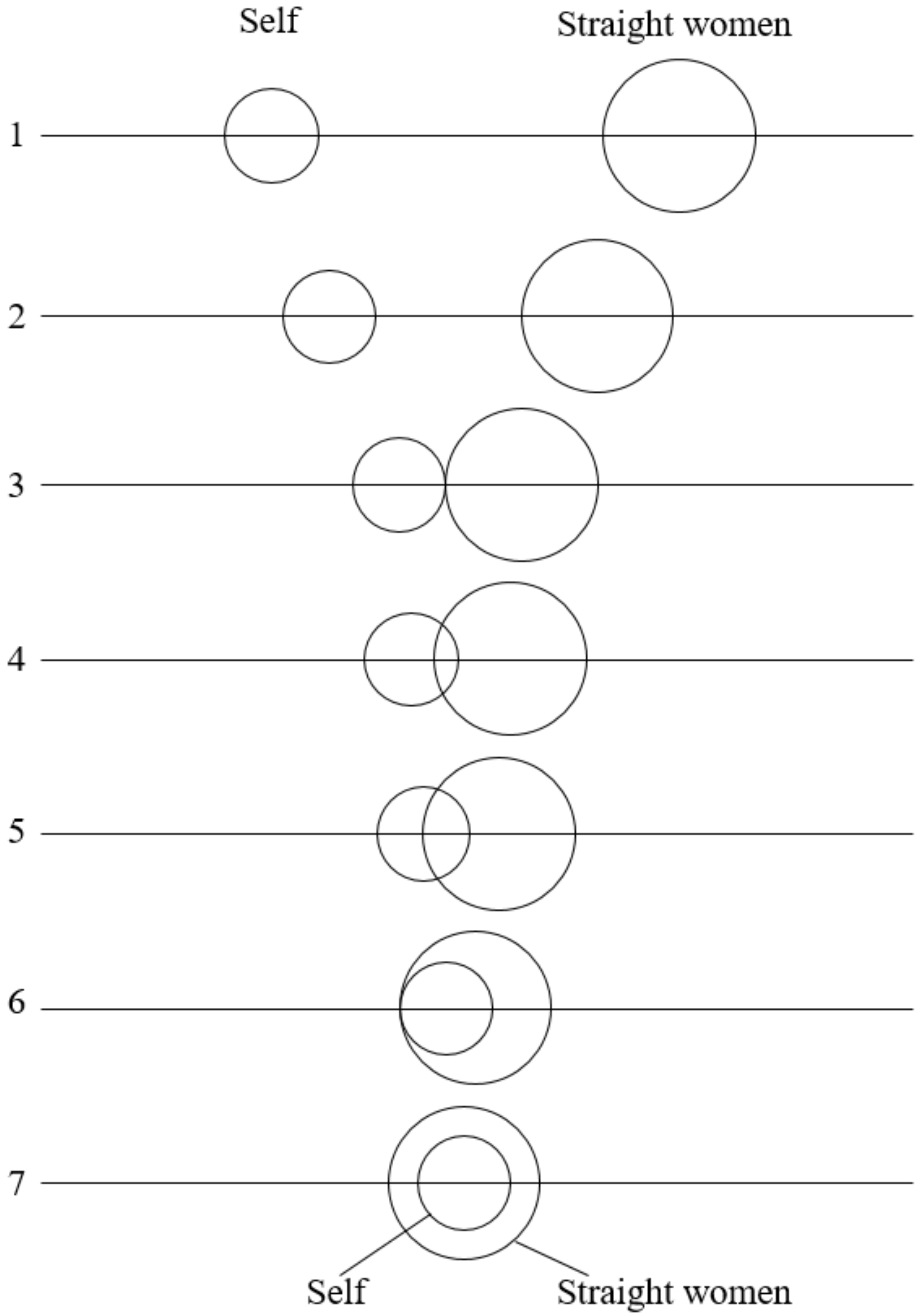
Group affiliation (Affiliation to lesbians)

Item: "Please mark the picture that describes your relationship to the group of lesbians best."



Group affiliation (Affiliation to straight women)

Item: “Please mark the picture that describes your relationship to the group of straight women best.”



Psycho-social identification

Collective Self-Esteem Scale (CSES by Luhtanen and Crocker, 1992).

	Strong rejection						Strong agreement
	1	*2*	*3*	*4*	*5*	*6*	*7*
Private-subscale							
*I often regret that I belong to this social group.	○	○	○	○	○	○	○
In general, I'm glad to be a member of the social group I belong to.	○	○	○	○	○	○	○
*Overall, I often feel that this social group is not worthwhile.	○	○	○	○	○	○	○
I feel good about the social group I belong to.	○	○	○	○	○	○	○
Identity-subscale							
*Overall, my group membership has very little to do with how I feel about myself.	○	○	○	○	○	○	○
The social group I belong to is an important reflection of who I am.	○	○	○	○	○	○	○
*The social group I belong to is unimportant to my sense of what kind of a person I am.	○	○	○	○	○	○	○
In general, belonging to this social group is an important part of my self-image.	○	○	○	○	○	○	○

*Items with asterisks had to be recoded (1 to 7, 2 to 6, 3 to 5, 4 to 4, 5 to 3, 6 to 2, 7 to 1).

Present social environment

Current contact to lesbians (by Herek, Gillis, Cogan, and Glunt, 1997).

Item: "Frequency of contact to lesbians."

1	2	3	4	5	6	7
Never	Almost never	Seldom	Sometimes	Often	Very often	Always

Current contact to straight women (by Herek, Gillis, Cogan, and Glunt, 1997).

Item: "Frequency of contact to straight women."

1	2	3	4	5	6	7
Never	Almost never	Seldom	Sometimes	Often	Very often	Always

Gender distribution of friends.

Item: "My circle of friends consists of..."

1	2	3	4	5	6	7
…men only	…men mainly	…somewhat more men	…men and women equally	…somewhat more women	…women mainly	…women only

Sexual orientation of female friends.

Item: "My female circle of friends consists of..."

1	2	3	4	5	6	7
… lesbians only	… lesbians mainly	…somewhat more lesbians	…lesbians and straight women equally	…somewhat more straight women	… straight women mainly	… straight women only

Childhood social environment (Study 2 only)

Contact to girls during childhood.

Item: "Contact to girls."

1	2	3	4	5	6	7
Never	Very seldom	Seldom	From time to time	Often	Very often	Always

Contact to boys during childhood.

Item: "Contact to boys."

1	2	3	4	5	6	7
Never	Very seldom	Seldom	From time to time	Often	Very often	Always

Coming-out measures (Data collected for non-straight women only)

Coming-out extent.

Item: "Does your mother know about your sexual orientation?" Yes/ no.

Item: "Does your father know about your sexual orientation?" Yes/ no.

(The scores for parents' items were adapted to the 7-point scale in order to get comparable values for an overall coming-out extent index: if neither mother nor father knew about their daughter's sexual orientation, this was recoded as "1"; if only one parent knew about it, it was represented with "4", and if both parents knew, a "7" was assigned.)

Item: "How many persons of the following groups, if existing, know about your sexual orientation?"

	None			Half of them			All of them
	1	*2*	*3*	*4*	*5*	*6*	*7*
Present straight friends	○	○	○	○	○	○	○
Acquaintances	○	○	○	○	○	○	○
Fellow students	○	○	○	○	○	○	○
Colleagues	○	○	○	○	○	○	○
Boss	○	○	○	○	○	○	○

Negative coming-out experiences.

	Strong rejection					Strong agreement
	1	*2*	*3*	*4*	*5*	*6*
Difficult Process(-subscale of LGIS by Mohr and Kendra, 2011).						
Admitting to myself that I'm an LGB person has been a very painful process.	○	○	○	○	○	○
Admitting to myself that I'm an LGB person has been a very slow process.	○	○	○	○	○	○
*I have felt comfortable with my sexual identity just about from the start.	○	○	○	○	○	○
Concealment Motivation(-subscale of LGIS by Mohr and Kendra, 2011)						
I keep careful control over who knows about my same-gender romantic relationships.	○	○	○	○	○	○
I prefer to keep my same-gender romantic relationships rather private.	○	○	○	○	○	○
My sexual orientation is a very personal and private matter.	○	○	○	○	○	○
Internalized Homonegativity(-subscale of LGIS by Mohr and Kendra, 2011).						
If it were possible, I would choose to be straight.	○	○	○	○	○	○
I wish I were straight.	○	○	○	○	○	○
I believe it is unfair that I am attracted to people of the same gender.	○	○	○	○	○	○

*Items with asterisks had to be recoded (1 to 6, 2 to 5, 3 to 4, 4 to 3, 5 to 2, 6 to 1).

Perceived discrimination experiences (Study 2 only).

	Never	Once	Rarely	Some- times	Often
	1	*2*	*3*	*4*	*5*
Other people act as if I was uncomfortable to them.	○	○	○	○	○
I was insulted or called names.	○	○	○	○	○
I feared something could happen to me.	○	○	○	○	○
Compared to other people I felt like I was treated unfair.	○	○	○	○	○
I was harassed.	○	○	○	○	○
I received poorer service in bars or restaurants.	○	○	○	○	○
In shops I was ignored or treated unfriendly.	○	○	○	○	○
I was threatened.	○	○	○	○	○
Other people treated me with less respect.	○	○	○	○	○
Other people ignored me.	○	○	○	○	○
I was treated less politely compared to other people.	○	○	○	○	○
I experienced disadvantages during renting a flat.	○	○	○	○	○
I lost my job or did not get one, because I’m lesbian or bisexual.	○	○	○	○	○

Manuscript 3

Kachel, S., Simpson, A. P., & Steffens, M. C. (2017).

Do I Sound Straight? – Acoustic Correlates of Actual and Perceived Sexual Orientation and Masculinity/Femininity in Men's Speech.

In press in *Journal of Speech, Language, and Hearing Research.*

“Do I Sound Straight?” –

Acoustic Correlates of Actual and Perceived Sexual Orientation and Masculinity/Femininity in Men’s Speech

Sven Kachel[a, b]

Adrian P. Simpson[a, c]

Melanie C. Steffens[a, b]

a) DFG Research Unit Person Perception, Institute of Psychology, Friedrich Schiller Universtity of Jena, Am Steiger 3/1, 07743 Jena, Germany
b) Department of Social, Environmental, and Economic Psychology, University of Koblenz-Landau, Fortstraße 7, 76829 Landau, Germany
c) Department of German Linguistics, Friedrich Schiller University Jena, Fürstengraben 30, 07743 Jena, Germany

Corresponding author: Sven Kachel (sven.kachel@uni-jena.de)

Version: 19th December 2017
In press in *Journal of Speech, Language, and Hearing Research*

ABSTRACT

Purpose: The present study aims to give an integrative answer on which speech stereotypes exist towards German gay and straight men, whether and how acoustic correlates of actual and perceived sexual orientation are connected, and how this relates to masculinity/femininity. Hence, it tests speech stereotype accuracy in the context of sexual orientation.

Method: 25 gay and 26 straight German speakers provided data for a fine-grained psychological self-assessment (e.g., masculinity/femininity) and explicit speech stereotypes. They were recorded for an extensive set of read and spontaneous speech samples using microphones and nasometry. Recordings were analyzed for a variety of acoustic parameters (e.g., fundamental frequency and nasalance). 74 listeners categorized speakers as gay or straight based on the same sentence.

Results: Most relevant explicitly expressed speech stereotypes encompass voice pitch, nasality, chromaticity, and smoothness. Demonstrating implicit stereotypes, speakers were perceived as sounding straighter, the lower their median f0, center of gravity in /s/, and mean F2. However, based on actual sexual orientation, straight men only showed lower mean F1 than gay men. Additionally, we found evidence that actual masculinity/femininity and the degree of sexual orientation were reflected in gay and straight men’s speech.

Conclusion: Implicit and explicit speech stereotypes about gay and straight men do not contain a kernel of truth and differences within groups are more important than differences between them.

Key words: intragroup differences, stereotypes, sexual orientation, masculinity, femininity, nasality

1. Introduction

"Do I sound straight?" is a question many men may have been confronted with when experiencing homo-negativity (such as verbal or physical abuse) – independent of their actual sexual orientation. As an illustration, a motion picture of almost the same title deals with the question when, why, and which gay men learn to sound gay and which straight men learn to sound straight, respectively (Gertler, Thorpe, & Thorpe, 2014). Through this film, a growing field in sociophonetic research has gained attention from a broader audience. However, the more fundamental point in this documentary movie is which speech stereotypes regarding straight and gay men do exist, whether they contain a kernel of truth, and how does this relate to masculinity/femininity. The aim of the present research is to investigate speech stereotype accuracy by using a large sample of German gay and straight men, analyzing a variety of acoustic parameters and relating these, in turn, to fine-grained psychological measures.

1.1 Systematization of previous studies

Stereotypes can be understood as associations of social groups (e.g., straight men) and certain features (e.g., low-pitched voice; see Allport, 1954). Related to sexual orientation, stereotypes are based on the assumption of gender (non)conformity. According to lay gender inversion theories (Kite & Deaux, 1987), lay people believe gay men compared to straight men to be more similar to straight women (effeminacy/de-masculinization of gay men; Madon, 1997). [1] Indeed, straight women and men have described themselves as more gender conforming than lesbians and gay men (meta-analysis by Lippa, 2005), and perceiving a person as straight is associated with perceiving her or him as gender conforming (Gaudio, 1994; Smyth, Jacobs, & Rogers, 2003; Camp, 2009). However, there is considerable intra-group variability regarding masculinity/femininity among lesbians, gay men, straight women, and straight men (Lippa, 2005;

[1] Implicitly following lay gender inversion theories, most previous studies on the acoustic correlates of sexual orientation investigated how gay men acoustically differed from straight men. By doing so, behavior that deviates from gender role norms is marked and gay men are made the effect to be explained (Bruckmüller, 2013). In contrast, the present paper takes another perspective by focusing on straight men's speech and how it differs from gay men's speech.

Kachel, Steffens, & Niedlich, 2016), and some evidence suggests that acoustic correlates of perceived sexual orientation and masculinity/femininity should be separated (Munson, 2007). Hence, sexual orientation and masculinity/femininity are not as interchangeable as lay people's gender inversion theories seem to suggest (see Kite & Deaux, 1987). We therefore treat acoustic correlates of sexual orientation and masculinity/femininity as related but not equivalent.

When looking at phonetic research on sexual orientation and masculinity/femininity, we can distinguish between speaker- and listener-centered studies. *Speaker-centered studies* focus on the speaker as the signal-producing actor within the communication process and ask which acoustic parameters a speaker uses to mark a certain identity or role. *Listener-centered studies* focus on the listener as the receptive actor within the communication process. They ask which acoustic parameters influence listeners' impressions of speakers as sounding lesbian/gay or straight and feminine or masculine, respectively.

In the following, we review speaker-centered studies, analyzing acoustic correlates of actual masculinity/femininity because we want to focus on the effects of speakers' social gender on speech (Smyth & Rogers, 2008; Kachel, Simpson, & Steffens, 2017). Hence, we are not concerned with *perceived* masculinity/femininity attributed to the speakers by a group of listeners (see e.g. Hancock, Colton, & Douglas, 2014). On this basis, we present different methods for determining explicit and implicit speech stereotypes. Subsequently, we focus on whether these speech stereotypes contain a kernel of truth by reviewing speaker-centered studies on acoustic correlates of sexual orientation that deal with natural recordings made on one occasion of a group of self-identified gay and straight speakers. Finally, we elaborate why these findings could be inconclusive.

1.2 Acoustic correlates of actual masculinity/femininity

Stereotypically, sexual orientation and masculinity/femininity are associated with each other (Kite & Deaux, 1987). Listener-centered studies have dominated research on the acoustics of masculinity/femininity and consistently showed that the more straight-sounding male voices were rated, the more gender conforming they were considered to be (Gaudio, 1994; Smyth et al., 2003;

Camp, 2009). Because the present study deals with actual instead of perceived masculinity/femininity, we review sociophonetic research investigating acoustic correlates of actual masculinity/femininity only. To our knowledge, there are only two identity-based speaker-centered studies that have dealt with actual masculinity/femininity in adults, one of them with $n = 2$ (Chiang, 2003; for actual masculinity/femininity in children see Li, Rendall, Vasey, Kinsman, Ward-Sutherland, & Diano, 2016). In the other study, 57 Dutch women and men each were asked to ascribe feminine and masculine personality traits and behaviors to themselves (Biemans, 2000). Except for the feminine behavior self-rating scale, utilized gender-related scales showed no gender differences. The disputable validity of the masculinity/femininity measurement used could account for the unexpected result patterns (e.g., men who self-described their behavior as more masculine showed a higher f0 range). In an attempt to remedy this, we will use the Traditional Masculinity-Femininity scale (Kachel et al., 2016) that has been shown to be a more reliable and valid measure of masculinity/femininity. By doing so, the present study contributes to a better understanding of acoustic correlates of actual masculinity/femininity.

1.3 Speech stereotypes and stereotype (in)accuracy

How can we test whether the stereotype of the low-pitched straight men contains a kernel of truth? Stereotype accuracy refers to the extent to which beliefs about social group characteristics correspond to their actual characteristics (Jussim, Crawdford, & Rubinstein, 2015). There are at least two different kinds of stereotype accuracy (Judd & Park, 1993): Stereotypic and dispersion accuracy. Stereotypic accuracy means that the extent of a stereotypical belief is a true reflection instead of an over- or underestimation of actual group characteristics with regard to content, e.g., the belief that straight men speak on a lower pitch than gay men mirrors the actual mean f0 difference between gay and straight men. Dispersion accuracy refers to a correspondence of the assumed and the actual group variability. Because stereotypes are generalized representations of social group characteristics (Judd & Park, 1993), members of the same group are believed to behave homogeneously while members of different groups are assumed to behave differently.

Consequently, dispersion accuracy would not be given if members of one group demonstrate varying characteristics because their dispersion would be underestimated, e.g., some straight men speak on a lower pitch while others speak at a higher pitch.

Identifying people's beliefs about the speech of gay and straight men is the first step in determining speech stereotype accuracy (Jussim et al., 2015). At least three approaches can be distinguished to provide information on speech stereotypes regarding sexual orientation. The first is inferring people's beliefs about the speech characteristics of gay and straight men from lay gender convergence theories. Following that approach, straight men are supposed to show more gender conforming speech compared to gay men. Given the small number of sociophonetic studies dealing with social gender, gender conforming speech patterns can be derived from studies comparing female and male speakers. Based on the finding that men speak with a lower-pitched voice than women (Munson & Babel, 2007; Simpson, 2009), straight men would be assumed to produce lower mean f0 than gay men.

Another method of determining speech stereotypes is directly asking lay people about their subjective impressions of straight versus gay men's speech. Based on open-ended interview questions, American English respondents described the speech of straight men in gender conforming terms: low pitch (Piccolo, 2008; Panfili, 2011), monotonous, bad articulation (Piccolo, 2008). By contrast, they typically characterized the speech of gay men as lisping, high-pitched (Piccolo, 2008; Panfili, 2011), variable in voice pitch, having a dramatic expression, and precise and clear articulation (Piccolo, 2008; see Mack, 2010, for similar findings in Puerto Rican-Spanish respondents). Until now, no quantitative approach has been used to explore speech stereotypes. Hence, in the present study we test stereotypes by using a list of voice and speech characteristics and focus on a German sample in order to test cross-cultural generalizability of speech stereotypes. Moreover, we are concerned with cultural specificity by analyzing nasality, which is one of the strongest speech stereotypes regarding male sexual orientation in German culture.

Both approaches – deductions from lay gender inversion theories and asking respondents –

can be used for determining explicit speech stereotypes (see Table 1). In contrast to explicit speech stereotypes, using indirect approaches does not rely on listeners' verbalizable knowledge and reduces social desirability biases. Conducting listener-centered studies is one way of indirectly ascertaining speech stereotypes by assessing which sexual orientation ratings by listeners accompany which acoustic parameters of speakers. Similar to other implicit measures, the indirect approach addresses implicit stereotypes by identifying acoustic correlates of perceived sexual orientation (Greenwald & Banaji, 1995). As shown in Table 1, findings for implicit stereotypes are inconclusive. For example, the prominent "gay lisp" stereotype (Piccolo, 2008; Mack, 2010; Panfili, 2011) was supported by two listener-centered studies that showed straight-sounding men to have lower measures of central tendency of /s/ than gay-sounding men (Linville, 1998; Sulpizio, Fasoli, Maass, Paladino, Vespignani, Eyssel, & Bentler, 2015: Italian men). However, three listener-centered studies did not find this difference (Munson, McDonald, DeBoe, & White, 2006; Zimman, 2010; Sulpizio et al., 2015: German men).

1.5 Kernel of truth: Actual acoustic correlates of sexual orientation

After determining speech stereotypes, the next step for assessing stereotype accuracy is identifying objectifiable criteria reflecting group characteristics (Jussim et al., 2015). Speaker-centered studies analyze whether speech stereotypes contain a kernel of truth by asking whether gay and straight men really differ in their speaking patterns. Actually, there is supporting evidence for the popular belief in different cultures that straight men can be distinguished from gay men based on their speaking patterns (Zwicky, 1997). Intonational features, spectral (e.g., sibilant measures, vowel space characteristics) and temporal information (e.g., segmental duration differences, voice onset time) as well as a few voice quality parameters have been investigated (e.g., creakiness, breathiness). Going beyond a corresponding review by Munson and Babel (2007), Table 1 gives a comprehensive overview of relevant studies on acoustic correlates of actual sexual orientation conducted within the last decade and relates them to explicit and implicit speech stereotypes. The main finding is that although a number of studies provide some evidence for explicit and implicit speech stereotypes being accurate, an equal or a larger number of studies showed no or even contradicting evidence. For example, this is true for the "gay lisp" stereotype: One study found actually straight men to produce lower measures of central tendency of /s/ than actually gay men (Linville, 1998), while another investigation did not (Zimman, 2010).

1.6 Shortcomings of previous sociophonetic studies on sexual orientation

Results of speaker- and listener-centered studies on sexual orientation are inconclusive for a number of reasons. Much earlier work was carried out in an American English-speaking context, while more recent studies on other Indo-European languages have provided more cross-linguistic data, e.g. Dutch (Baeck, Corthals, & van Borsel, 2011), Czech (Valentova & Havlíček, 2013), Canadian English (e.g., Rogers & Smyth, 2003), Canadian French (Sisson, 2003), Italian, and German (Sulpizio et al., 2015). However, even when looking at previous studies referring to the same language, such as American English, the pattern of results is inconsistent. This inconsistency can be due to the same reasons mentioned in a recent overview for female speakers (Kachel et al., 2017). First, different speech materials have been used, ranging from spontaneous speech (Smyth et al., 2003) to groups of words (Munson et al., 2006) which is accompanied by divergent acoustical analyses. For example, inconsistent findings on intonational differences have been partly attributed to measurements at vowel midpoints of single words vs. measurements on connected read speech (Baeck et al., 2011). Second, the test power could be improved by increasing the number of speakers per group. Less than half of the previous studies dealt with speaker sample sizes of more than 13 men per group (e.g., Rendall, Vasey, & McKenzie, 2008). Third, different definitions, operationalization, and classification strategies of sexual orientation have been applied. While some studies compared openly gay men with men not open about their sexual orientation (Lerman & Damsté, 1969; Gaudio, 1994), others contrasted non-straight and straight men (e.g., Munson et al., 2006; Rendall et al., 2008) or gay and straight men (e.g., Pierrehumbert, Bent, Munson, Bradlow, & Bailey, 2004). Self-categorization measures give a false impression of a clear division between sexual orientation groups and suggest homogeneity within groups (see Kachel et al., 2017). Hence, utilizing the common more fine-grained Kinsey-scale (Kinsey, Pomeroy, & Martin, 1948) can help to overcome these problems because it captures variance in sexual orientation within a sexual orientation group (e.g., Linville, 1998) and exposes intra-group differences (Kachel et al., 2017). Finally, previous sociophonetic studies on sexual orientation almost entirely missed collecting data on additional speaker characteristics possibly explaining speech differences due to sexual orientation. One of the few exceptions is a study by Rendall and colleagues (2008), who considered physical measures and showed that differences in vowel-specific F2-values between gay and straight men were caused by body size differences. Next to physical attributes, knowledge on psychological characteristics influencing sexual orientation differences in speech is similarly sparse.

1.7 Psychological characteristics related to sexual orientation

In sociophonetic research, psychological speaker characteristics are assumed to provide an answer to the question which straight men perform "straight speech" and which gay men perform "gay speech". However, only a small number of studies considered psychological characteristics. In line with Kachel et al. (2017) who showed dispersion inaccuracy for female sexual orientation groups, the present study tests the roles of masculinity/femininity, group affiliation, psychosocial identification, social environment, and coming-out measures for male speakers.

Masculinity/femininity has been discussed as an important psychological speaker variable illuminating the acoustic correlates of sexual orientation in early research (Lerman & Damsté, 1969). However, no previous sample-based studies have analyzed it with regard to sexual orientation (for a case study see Chiang, 2003). We expect gay and straight men who conform to a more masculine gender role to produce more gender conforming speech.

Gay-sounding speech is assumed to function as a marker of *group affiliation* – the degree to which a person describes her- or himself as a group member – to the gay community (Linville, 1998). Similarly, gay men showing a high psychosocial identification with the ingroup (i.e., whose group membership is an important and satisfying part of their self-concept), were hypothesized to acquire gay speech by identifying with gay adult role models and imitating their speech (Pierrehumbert, et al., 2004; Rendall et al., 2008). Hence, we expect straight men who are highly affiliated and identify with their in-group to produce gender conforming speech (vice versa for gay men).

The role model acquisition scenario emphasizes the relevance of the *social environment* (Smyth & Rogers, 2008). Gay men may not feel affiliated to other gay men (vice versa for straight men), nor may they have contact to them. In addition to adulthood, childhood has been assumed to be a critical phase in adopting speaking patterns associated with sexual orientation (Rendall et al., 2008; Smyth & Rogers, 2008). Hence, men reporting more contact to straight than gay men and more contact to boys than girls during childhood should show more gender conforming speech.

Finally, within-group variability regarding gay men's *coming-out experiences* can possibly influence their speech (Jacobs, 1996; Zwicky, 1997). Although gay men's coming-out age and extent was not related to mean f0 (Baeck et al., 2011), it is unclear whether those who feel uncomfortable about coming out use other acoustic parameters in a gender conforming manner.

1.8 The present study

The present study takes an integrative perspective on explicit and implicit speech stereotypes regarding sexual orientation, acoustic differences between and within gay and straight men, and their relations to actual masculinity/femininity. By doing so, the present study aims to provide empirical evidence to the superordinate question: How accurate are speech stereotypes regarding male sexual orientation? In order to answer this question, we recorded more than 50 German men almost evenly distributed across gay and straight speakers and collected perceived sexual orientation ratings from a large listener sample. We analyzed an extensive range of acoustic parameters including nasalance, a comparatively simple measure of analyzing nasality, which is one of the most important speech parameters associated with male sexual orientation in German speakers. Moreover, we provide a fine-grained analysis of actual sexual orientation. The most innovative aspect of our study is that we collected data on a variety of psychological characteristics in order to represent intra-group variability, including a valid and reliable measurement of actual masculinity/femininity (Kachel et al., 2016). We ask the following questions:

1) *Actual masculinity/femininity:* What are the acoustic correlates of men's actual masculinity/femininity?
2) *Explicit speech stereotypes:* Which speech stereotypes are mentioned regarding straight and gay men?
3) *Implicit speech stereotypes:* Which acoustic parameters influence sexual orientation ratings? Do explicit speech stereotypes correspond to implicit ones?
4) *Actual sexual orientation – gay vs. straight men:* Do gay and straight men acoustically differ in German? Can the findings from a smaller German speaker sample be replicated (Sulpizio et al., 2015)? Do explicit and implicit speech stereotypes contain a kernel of truth?
5) *Actual sexual orientation – within-group variability:* Are there any acoustic differences between mainly and exclusively gay/straight men? Do psychological characteristics influence gay men's speech?

2. Method

2.1 Participants

Data were gathered from 54 male speakers living in and around the east German town of Jena. Speakers were invited to take part in an investigation on "Sexuality and language" via bulletin boards, handouts, and mailing lists at the university, as well as directly through LGBTIQ* organizations and platforms. Using a snowball-technique, participants were asked to encourage friends and acquaintances to participate. In order to reach an adequate number of participants, the study description explicitly contained a request for men divergent in sexual orientation. Only German native speakers without voice and speaking disorders were considered.

Speakers were first asked to complete a questionnaire eliciting detailed information on sexual orientation, masculinity/femininity, and related psychological characteristics. Sexual orientation information was used to oversample the gay population. The sample is highly homogeneous with regard to ethnicity, educational level, and age. All men were Caucasian, ranged in age from 20 to 30 years (M = 24.52, SD =2.53), and nearly all studied at the local university. According to the 7-point Kinsey-like scale (1-exclusively gay, 4-equally gay and straight, 7-exclusively straight; modified version from Kinsey et al., 1948), 25 men identified as gay (Kinsey-like score ≤ 2) and 26 as straight (Kinsey-like score ≥ 6). In order to check for group variability within sexual orientations, we contrasted mainly vs. exclusively gay and straight men, respectively. Three speakers with Kinsey-like scores of 3–5 were excluded from further analyses (see Table A1 in the Online Appendix for the detailed distribution along the Kinsey-like scale). Gay and straight men did not differ significantly on sociodemographic characteristics such as age (M_{gay} = 24.28, $M_{straight}$ = 24.85; $t[48] = -0.78$, $p = .44$).

2.2 Psychological characteristics

The complete online questionnaire is available at (Online Appendix). All internal consistencies of the described scales were Cronbach's $\alpha \geq .73$, indicating reliable measurement.

2.2.1 Sexual orientation (Gynophily-androphily difference)

In addition to sexual self-identification, we used more objectifiable questions in order to measure sexual orientation (Worthington & Reynolds, 2009; also see Klein et al., 1985; Shively, Jones, & DeCecco, 1985). Sexual orientation towards women and men was measured separately with four items each: sexual fantasies, romantic emotions, physical attraction, and sexual interaction (e.g., "Physical attraction towards women"). These were rated in order to answer the question "How often have you experienced..." on a 7-point scale ("1= never", "7 = always"). Due to the high negative correlation between sexual orientation towards women and men ($r = -.93$, $p < .001$), a gynophily-androphily difference was computed for each speaker. The higher values are above zero, the straighter a speaker is. Conversely, values below zero indicate an increasing attraction towards men. Values around zero indicate a more bisexual orientation. The gynophily-androphily difference and the Kinsey-like scale were highly correlated ($r = .98$, $p < .001$).

2.2.2 Masculinity/femininity

Three questionnaires were used for measuring gender-role self-concept.

a. TMF

First, speakers rated themselves on the 7-point (“1 = very masculine”, “7 = very feminine”) Traditional Masculinity-Femininity Scale (TMF; Kachel et al., 2016), which consists of six bipolar items such as “I consider myself…” and “Traditionally, my behavior would be regarded as…”. Higher averaged scores indicate more femininity/less masculinity.

b. GEPAQ

Speakers completed the German version (GEPAQ; Runge, Frey, Gollwitzer, Helmreich, & Spence, 1981) of the Extended Personal Attributes Questionnaire (Spence & Helmreich, 1978). This is a specific, personality-based measure of present gender-role self-concept, with two scales using eight items each. On the first scale (GEPAQ-M), instrumental, goal-oriented traits, typically associated with men, like independence and self-confidence, were assessed on a 5-point scale. The second scale (GEPAQ-F), assesses stereotypical female traits, more oriented toward social relations and related to emotional expressiveness, such as helpfulness and kindness. Higher average scores indicate a higher degree of masculinity on GEPAQ-M and of femininity on GEPAQ-F.

c. Gender-role self-concept during childhood

The Childhood Gender Nonconformity Scale (Rieger, Linsenmeier, Gygax, & Bailey, 2008) was used to assess how well a speaker considered himself to have rejected common gender stereotypes during childhood. It consisted of seven statements like “As a child I preferred playing with girls rather than boys”. In contrast to the original instrument, we chose a 5- instead of a 7-point scale (“1 = strong disagreement”, “5 = strong agreement”), modified some wordings, and added the item “As a child I behaved like a typical boy”. Higher scores indicated a higher degree of gender nonconformity during childhood.

2.2.3. Group affiliation

Two circle items (see Schubert & Otten, 2002) were used to separately measure the degree of self-assigned affiliation to gay and straight men. Each item consisted of seven graphics visualizing the distance between a smaller circle (self) and a larger circle (group). The smaller the distance and the higher the overlap between the two circles, the higher the degree of group affiliation (see Online Appendix for visualization).

2.2.4. Psychosocial identification

Two word-based scales were taken from Luhtanen and Crocker's (1992) Collective Self-Esteem Scale in order to assess speakers' social identity as gay or straight. Both measures contained four items that were rated on a 7-point scale ("1 = strong rejection", "7 = strong agreement"). The "Identity"-subscale was used to determine the group membership's importance (e.g., "The social group I belong to is an important reflection of who I am"). The "Private"-subscale assessed speakers' satisfaction with group membership (e.g., "In general, I'm glad to be a member of the social group I belong to"). Because of the overall scale's high reliability (Cronbach's α = .85), we created an overall index, higher scores indicating higher identification with the in-group.

2.2.5. Social environment

a. Present social environment.

Based on Hodson, Harry, and Mitchell (2009) speakers evaluated their current contact to gay and straight men on two separate 7-point scales ("1 = never", "7 = always"). Data for general gender distribution of friends and sexual orientation of male friends were also collected on two 7-point scales ("1 = women/gay men only", "7 = men/straight men only"). Higher scores indicated a higher proportion of male and straight male friends.

b. Childhood social environment.

Similar to current contact to gay and straight men, we assessed contact to girls and boys during childhood. We used two 7-point frequency-scales ("1 = never", "7 = always"), one for contact to girls and one for contact to boys. Higher scores indicated more frequent contact to girls/boys during childhood.

2.2.6. Voice self-assessment

Data on voice self-assessment were collected during the second measurement. It encompassed six items, such as "I think I speak stereotypically straight" rated on a 7-point scale ("1

= strong rejection”, “7 = strong agreement”). Higher scores indicate a more gender conforming assessment.

2.2.7 Coming-out measures

Three measures related to speakers’ coming-out were collected from gay men only.

a. Coming-out extent

Coming-out extent represents the visibility of same-gender orientation to the social environment as well as negative experiences during the coming-out process. Speakers rated parental knowledge of their gay orientation for mother and father separately. Subsequently, they provided information on how many persons of different groups (e.g., colleagues) know about their sexual orientation (“1 = none, “7 = all of them”). To obtain comparable values for the parents’ questions, the scores were adapted to the 7-point scale: If neither mother nor father knew about the sexual orientation, this was recoded as “1”; if only one parent knew, it was represented with “4”, and if both parents knew, “7” was assigned. Higher average scores indicate a higher coming-out extent.

b. Concealment motivation

One subscale of the Lesbian, Gay, and Bisexual Identity Scale (Mohr & Kendra, 2011) was used to examine preference to hide one’s sexual orientation. It consisted of three items like “I keep careful control over who knows about my same-gender romantic relationships” rated on a 6-point scale (“1 = strong rejection”, “6 = strong agreement”). In contrast to the original wording, we used “same-gender” instead of “same-sex”. Higher scores indicated a higher concealment motivation.

c. Perceived discrimination experiences

To measure experience of discrimination, two scales were used. The first is based on Steffens, Bergert, and Heinecke (2010), which is a translated and modified version of Herek and Berrill’s (1990) Sample Survey of Anti-Gay Violence and Victimization. Speakers were asked about concrete events, such as “people threatened or insulted me”. The second is the Everyday Discrimination Scale (Williams, Yu, Jackson, & Anderson, 1997), originally developed to assess soft forms of discrimination regarding race, like “shop assistants treated me with poorer service”.

This scale was adapted to sexual orientation and restricted to experiences within the last five years ("1 = never", "5 = often"). Higher scores indicated a higher degree of perceived discrimination.

2.2.8 Explicit speech stereotypes

In order to measure which voice and speech characteristics are typically associated with straight and gay men, we created a list encompassing 40 items for lay persons. Hence, the formulation of these items is more impressionistic (e.g., "rather asking") than acoustically correct (e.g., "rising f0 movement at the end of an utterance"). Among them, 32 denoted one end of 16 bipolar scales (e.g., "high" vs. "low", "nasal" vs. "non-nasal") while the other eight characteristics formed a unipolar scale each (e.g., "hoarse", "shrill"). Each item's association with straightness and gayness was assessed on a 7-point scale ("1 = gay only", "7 = straight only"). Higher scores indicated voice and speech features more associated with straightness and less with gayness.

2.3 Speech recordings

Recordings were made on two separate occasions. In the first session, conventional microphone recordings were made. In a second session, some months later, a subset of the same speaker group was recorded using nasometry to investigate possible differences in nasalance.

2.3.1 First session

Speakers were recorded in a sound-treated room using a condenser microphone (AKG C1000S) connected to an external pre-amplifying audio interface (M-Audio Fast Track Pro) at a sampling rate of 44100 Hz and 16-bit amplitude resolution. In order to prevent overmodulation when articulating plosives, the microphone was not located directly in front of the speakers but was placed at an angle of 45°/30° on the horizontal/vertical plane to the speakers. Moreover, the speakers were asked to maintain a mouth-microphone distance of approximately 10–15 cm. The recording was started by asking the speakers to count aloud from 15 to 25 in a slow and monotonous voice. Afterwards, the experimenter asked only those speakers to speak in their habitual voice who he did not believe to have done so while counting.

An extensive set of read and spontaneous speech samples were elicited. Spontaneous speech included the descriptions of a simple drawing of a farmyard and an approximately 500-meter route through town, while the read speech comprised one short text, two sets of individual sentences, and isolated vowels. The short text ("The butter story") is well-established elicitation text that allows for comparison and integration with other existing corpora (e.g. *Kiel Corpus of Read Speech*, Simpson, Kohler, & Rettstadt 1997). Two sentence sets were used to explicitly elicit particular features. In contrast to the first one containing ten sentences, the second set has already been used to elicit data in a study of gender-related vowel differences (Weirich & Simpson, 2013). Included in a list of twenty sentences are three sentences shown in (1), containing the abbreviations BII, LUU, IAA. These abbreviations elicit sequences of long point vowel tokens /i: a: u:/ with the stress on the final vowel. We can expect vowels in such sequences to be minimally affected by any neighboring consonants or temporal constraints that might give rise to undershoot.

(1)

Sie fuhren letzte Woche zur IAA nach Frankfurt. ("They went to the IAA in Frankfurt last week."); *Sie fahren nächste Woche zur LUU nach Hannover.* ("They're going to the LUU in Hannover next week."); *Wir wollen am Wochenende zur BII nach Hamburg.* ("We want to go to the BII in Hamburg at the weekend.")

In most varieties of German, including the Standard, we would also expect speakers to exhibit a certain amount of junctural glottalization at vowel onset, often taking the form of creaky voice (Kohler, 1994). The low fundamental frequency typically associated with different types of glottalization ensures tight interharmonic spacing, in turn allowing a more reliable formant estimation, especially in otherwise higher pitched female voices (Weirich & Simpson, 2013). Isolated vowels were elicited by asking speakers to utter the German letters <*a e i o u ä ö ü*> at normal citation rate, then <*a i u*> each for as long as possible. The long utterance of *a* was used to measure various voice quality parameters (see below). At the end of the first speech recording session, data on body measures were collected (see Rendall et al., 2008). Body weight was assessed in kilograms by using a standard bathroom scale and body height was collected in centimeters by measuring our speakers standing straight against a wall using a folding rule. Gay and straight men neither differed in body weight (M_{gay} = 77.10, $M_{straight}$ = 82.22; t[49] = –0.59, p = .56) nor height (M_{gay} = 181.76, $M_{straight}$ = 182.79; t[49] = –1.22, p = .23). Hence, weight and height were not taken into account in further analyses.

2.3.2 Nasometry session

The stereotypical "gay nasal" in German is equivalent to hyponasality[2], rather than hypernasality. To reliably measure this, nasometry (NasalView, Tiger DRS Inc., Seattle) was used (Fletcher, 1970). The acoustic output of the nose and the mouth is captured using two microphones fixed on the nasal and oral side of an attenuating plate that assures an acoustic separation of the two signals of approximately 20 dB (see Figure A1 in the Online Appendix). The intensity of the nasal signal in relation to the summed intensities of the nasal and oral signals represents the nasalance score in percent. Several months after the first recording session, 39 of the original speakers were recorded, maintaining a similar distribution with respect to self-reported sexual orientation: 19 gay, 18 straight (and two bisexual speakers who were excluded from analysis). Speakers were recorded producing a subset of the elicitation material (spontaneous speech: picture description; read speech: short story, two short sentence lists). The two sentence sets were read twice each.

2.4 Analysis

Temporal and spectral analyses were carried out on different parts of the elicited data sets to investigate acoustic features that have been found to vary between groups differing in gender and sexual orientation: intonational characteristics (fundamental frequency average and range), sibilant measures (spectral structure of the sibilants /s/ and /ʃ/), acoustic vowel space features, plosive characteristics (voice onset time and aspiration), and voice quality parameters including nasalance.

---

[2] This is illustrated well by the German comedian Hape Kerkeling in the following clip: https://www.youtube.com/watch?v=RjrnlzXKBrA

2.4.1 Intonational characteristics

Each speaker's average fundamental frequency and range was calculated for read and spontaneous speech. A *praat* script calculated the median together with the difference between the 2.5 and 97.5 percentiles as a measure of f0 range from one read task ("Butter story") and two spontaneous tasks (path description and picture description). Fundamental frequency was calculated using *praat*'s "To pitch…" routine with standard settings (floor: 75 Hz, ceiling: 600 Hz).

2.4.2 Sibilant measures

Sibilants, and in particular /s/, have been shown to be used by speakers to indexicalize group identity (e.g., Stuart-Smith, 2007). Spectral characteristics of both of the German fortis sibilants /s/ and /ʃ/ were measured, on the one hand, to establish if there are any intra- and intergroup differences in the speaker sample for individual sibilants, but also to investigate any differences in the size of the acoustic contrast between /s/ and /ʃ/. Using a *praat* script (Boersma, 2001), each sibilant was extracted, resampled at 22050 Hz and band-pass filtered to remove low-frequency energy below 200 Hz. The two spectral moments, center of gravity (CoG) and skewness, were calculated from a spectrum of the whole sibilant segment to reflect how energy is distributed in the spectrum. Furthermore, as a measure of contrast size, the difference between the CoG for /s/ and /ʃ/ for each speaker was calculated. Rather than using all of a speaker's sibilant tokens, two nouns/adjectives were chosen for each sibilant: /s/ from *Nässe* ("wet") and *mies* ("nasty"), /ʃ/ from *Schaufenster* ("shop window") and *neidische* ("envious"). Means per sibilant and speaker were calculated from the three repetitions of each word for each speaker.

2.4.3 Vowel space features

Estimates of the first and second formant frequencies of the point vowels /iː aː uː/ in vowel sequences contained in the abbreviations *BII, IAA, LUU* were used to define a speaker's acoustic vowel space. A *praat* script was used to perform a formant analysis on a stretch of the vowel sequence extending from 50-70% of the total duration of the whole sequence. The relevant stretch was extracted and downsampled to 6 kHz before performing formant analysis (burg) with the following settings: time step: 0.01 s; maximum number of formants: (3 for /iː aː/ or 4 for /uː/); maximum formant: 3000 Hz; window length: 0.025 s; preemphasis: 50 Hz). Modifying the maximum number of formants to be found to 4 for /uː/ significantly increased the reliability of the algorithm to separate F1 and F2 which are very close together for this German vowel, especially in a context of weak coarticulation (see above) where tongue-backness and lip-rounding are maximized, in turn giving rise to very low F2 close to F1. Average F1 and F2 values were calculated and manually checked for each vowel stretch. The area of the F2xF1 /iː aː uː/ triangle was calculated for each speaker and represented in kHz^2. Means per vowel and speaker were calculated from the three repetitions of each abbreviation for each speaker.

2.4.4 Plosive characteristics

The six German lenis and fortis plosives /p–b, t–d, k–g/ occurred in accented syllable onset of sentence-final past participle verb forms (*gebacken* "baked", *gepachtet* "rented out", *gedauert* "lasted", *getaucht* "dived", *gegart* "cooked", *gekauft* "bought") in six different sentences in one of the sentence sets, e.g. *Wie lang hat es denn gedauert?* ("How long did it last then?"). In accented syllable onset, fortis plosives /p t k/ are voiceless and aspirated. By contrast, the lenis series is voiced or devoiced and has a short voice onset time following plosive release. In its simplest form aspiration is coextensive with VOT. However, for a number of speakers, the onset of voicing after plosive release is still markedly breathy, correlating spectrally with predominant noise in the upper spectrum. This breathy voiced stretch (annotated here with VOT+) was delimited visually with the clear onset of F2. (Two typical examples considered to be with and without a period of breathy voice following VOT are shown in Figure A2 in the Online Appendix). Means per plosive and speaker were calculated from the three repetitions of each word for each speaker.

2.4.5 Voice quality parameters

Common voice quality parameters were analyzed from long, isolated /aː/ tokens. Jitter (ppq5), shimmer (apq5) and harmonics-to-noise ratio were determined for the whole vowel stretch using *praat*'s *voice report* facility. Moreover, average nasalance values for each reading and spontaneous task were calculated using NasalView's proprietary software. Averages for each task were used to calculate a single mean for each speaker.

2.5 Perceived sexual orientation

In addition to self-reported sexual orientation (i.e., Kinsey-like scores) and explicit speech stereotypes, we gathered data on perceived sexual orientation. A group of 74 listeners rated the straightness of each speaker's voice in an online listening study. Data from four raters were excluded (because of hearing impairments, being a non-native German speaker, previous participation in a study on the perception of sexual orientation, or missing participant code). Hence, data from 70 people were analyzed (53 female, 16 male, 1 other, $M_{age} = 23.13$, $SD_{age} = 3.67$, age range 18 – 43 years). The study was advertised via a students-only mailing list at the University of Koblenz-Landau as investigating the belief that people's sexual orientation could be recognized based on their voices. A single utterance produced by each speaker of one sentence from the read speech data from the first recording session *Der Tag ist sehr lang geworden.* ("It has been quite a long day.") was played via headphones to the raters, who were instructed to categorize the speaker's straightness as gay or straight. Female and male raters did not differ significantly in their ratings of gay and straight speakers (both $M_{Diff} \leq .01$, both ts[67] $\leq .40$, ps $\geq .69$). Hence, a mean perceived straightness score was calculated for each speaker across raters. The straighter speakers described themselves on the Kinsey-like scale, the more likely they were to be perceived as straight by raters ($r = .29$, $p = .037$). We refer to these perceptions as implying implicit speech stereotypes. Thus, explicit speech stereotypes (see scale described above) could be compared with these implicit speech stereotypes and with actual speech differences (speakers' acoustic characteristics) depending on sexual orientation.

3. Results

We first examined whether gay and straight speakers differed in any of the 16 psychological characteristics collected (see Table 2) and adjusted the significance level to multiple tests using the Bonferroni correction formula ($\alpha = .0031$). Regarding gender-role self-concept, straight men described themselves to be more masculine on TMF and to be more gender conforming during childhood than gay men did. In terms of group affiliation, straight speakers were more strongly affiliated to the straight group than gay men, whereas gay men were more strongly affiliated to the gay group than straight men. Regarding social environment, by comparison to gay men, straight men reported significantly higher current contact to other straight men but less current contact to gay men. Moreover, in contrast to gay men, straight men reported more often having more male friends who were straight. Finally, straight men showed higher contact to boys but less contact to girls during childhood than gay men did. Regarding voice self-assessment, straight men exhibited more gender conforming impressions of their own voices compared to gay men. Taken together, we found expected differences in psychological characteristics between gay and straight men.

3.1 Acoustic correlates of present gender-role self-concept

Are there acoustic correlates of present gender-role self-concept in men? In our sample consisting of males only (i.e., holding biological gender constant), we collected data on masculinity/femininity. Because TMF was shown to be better at differentiating gay and straight men compared to adjective-based measures of gender-role self-concept such as GEPAQ-M (Kachel et al., 2016), we calculated correlations for TMF only (adjusted significance level: $\alpha = .0018$, see Table 3). Men who claimed to be more feminine on TMF showed higher mean F2 values ($r = .49$, $p < .001$) and higher CoG values in /s/ ($r = .46$, $p = .001$, see Figure 1).

However, these acoustic correlates of masculinity/femininity could be mere artifacts of sexual orientation, because straight men described themselves as more masculine than gay men did. To test this, we computed correlations for gay and straight men separately. For mean F2 the association was found for both subgroups: The more feminine gay and straight men described themselves, the higher their mean F2 (both *r*s ≥ .39, both *p*s ≤ .046). For CoG in /s/ the association was not statistically significant (both *r*s ≤ .38, both *p*s ≥ .057). Thus, the association of gender-role self-concept and center of gravity in /s/ found for the overall sample is an artifact of actual sexual orientation because straight men who were more gender conforming produced lower centers of gravity in /s/ than gay men who were less gender conforming and produced higher centers of gravity in /s/. Instead, men who self-reported to be gender conforming showed more gender conforming mean F2 values that were not mere artifacts of sexual orientation, in line with lay people's gender inversion theory.

3.2 Explicit speech stereotypes

In order to investigate explicit speech stereotypes regarding straight and gay men, we examined whether the 40 voice and speech characteristics differed significantly from the scale midpoint using one-sample t-tests (test value = 4, adjusted significance level: $\alpha = .00125$). As Figure 2 shows, 24 voice and speech characteristics significantly differed from the scale midpoint. The most pronounced speech stereotypes (i.e., at least ± 1 from scale midpoint) in their order of magnitude were "low", "strong", "hard", "non-nasal", and "dark" for straight men and "nasal", "high", "soft", "shrill", and "bright" for gay men. The differences of the mean scores of "low" and "high" ($\Delta M_{low\_high} = 2.65$) as well as "non-nasal" and "nasal" ($\Delta M_{non\text{-}nasal\_nasal} = 2.59$) were similarly large ($|t[48]| = .23$, $p = .81$). Hence, nasality is as important as voice pitch when stereotypically characterizing the speech of German straight and gay men.

3.3 Implicit speech stereotypes

Determining the acoustic correlates of perceived straightness (adjusted significance level: $\alpha = .0018$), we found that the more men expressed gender conforming speaking patterns, the more likely they were to be perceived as straight. This was true for F2 in /a:/, mean F1 and F2, center of gravity in /s/, and median f0 in spontaneous speech (all $|r| \geq .44$, $p \leq .001$).

In order to quantify their relative importance, we used a multiple linear regression model with stepwise inclusion of the five predictors (see Table 4). The model explained 42% of the variance in perceived sexual orientation. Men were perceived as straighter, the lower their median f0 in spontaneous speech (β = -.35, p = .003), the lower their center of gravity in /s/ (β = -.33, p = .005), and the lower their mean F2 (β = -.29, p = .013). F2 in /aː/ and mean F1 seem to contribute redundant information and were not included in the regression model (both $\beta \leq$ -.12, both ps $\geq$ .58).

3.4 Acoustic differences between gay and straight speakers

Differences between gay and straight men were examined regarding 28 acoustic parameters (see Table 3, adjusted significance level: α = .0018). Straight men were significantly lower in F1 in /aː/ and mean F1 than gay men (both $|ts[49]| \geq 3.48$, ps $\leq$ .001; all other $|ts[49]| \leq 2.74$, ps $\geq$ .009). Although F1 in /aː/ was used to calculate mean F1 and therefore both parameters correlated highly (r = .82, $p \leq$ .001), mean F1 seems to contain additional information going beyond that of F1 in /aː/ (R^2 = .67). Hence, we estimated the relative importance of both acoustic parameters in explaining actual sexual orientation self-categorization as gay or straight by using a binary logistic regression model using the enter-method. The model explained 40% of the variance (Nagelkerke's R^2 = .40). Mean F1 (B = -.07, SE = .03, p = .022) predicted actual sexual orientation categorization significantly, whereas F1 in /aː/ did not (B = .00, SE = .01, p = .97). Hence, mean F1 was a much better predictor of actual sexual orientation categorization compared to F1 in /aː/.

3.5 Differences within the subgroups of gay and straight men

We next considered whether it is appropriate to treat gay (Kinsey-like scores 1-2) and straight men (Kinsey-like scores 6-7) as acoustically homogeneous entities. Because no previous study has investigated the acoustic correlates of sexual self-identification *within* gay and straight men (Baeck et al., 2011 had done so for gay men only), the significance level was left unadjusted for exploratory reasons (α = .05). As can be seen in Table 3, one acoustic parameter significantly differed each within the gay and the straight group. Exclusively gay men had less jitter in /aː/ than mainly gay men (t[23] = -2.22, p = .037) and exclusively straight men had a lower CoG in /s/ than mainly straight men ($t[24] = 2.11$, $p = .045$).

3.6 Correlations of psychological characteristics and acoustic measures within the gay group

Do psychological characteristics influence gay men's speech? Again, the significance level was left unadjusted for exploratory reasons, but to reduce the number of erroneously significant correlations, only those psychological features with at least three acoustic correlates were considered (see Table 5). Nearly all psychological features gathered were included in this analysis except for TMF (see above). Two out of 5 psychological characteristics that showed three or more acoustic correlates referred to gender-role self-concept (GEPAQ-F, CGNS). The more feminine gay men described themselves on GEPAQ-F, the higher their F1 and F2 in /uː/, CoG in /ʃ s/, VOT in /t/, and aspiration in /k/ ($|r| = .40$, $p < .048$). This pattern is predicted by gender inversion theory and occurs for acoustic correlates of CGNS as well: gender non-conform self-descriptions were accompanied by gender non-conform speaking patterns.

Contact to girls during childhood is the only psychological characteristic referring to social environment. The more contact to girls during childhood gay men reported, the higher their f0 range in spontaneous speech, the lower their shimmer in /aː/, and the lower their nasalance ($|r| = .42$, $p = .036$). Regarding voice self-assessment, the lower mean F1, median f0 in spontaneous speech, and the higher their HNR in /aː/, the more likely men described their voices as gender conforming and heteronormative. Hence, gender conforming voice self-assessment was accompanied by gender conforming speaking patterns. Discrimination experience was the only psychological characteristic referring to coming-out measures. Moreover, it was the only psychological characteristic showing acoustic correlates referring to one class of acoustic parameters: vowel space characteristics. Gay men exhibiting higher F1 in /iː/, F2 in /uː/, and mean F2, reported more discrimination experiences.

Summing up, each class of psychological characteristics, except for group affiliation measures, was represented acoustically and findings were in line with lay people's gender inversion theory. Hence, gay men showing gender conforming psychological characteristics exhibited gender conforming speaking patterns, and vice versa for gender non-conform gay men.

4. Discussion

How accurate are speech stereotypes regarding male sexual orientation? The present investigation represents the first study dealing with explicit and implicit speech stereotypes, acoustic differences between and within gay and straight men, and effects of social gender on speech by applying a fine-grained measurement of psychological characteristics and analyzing a wealth of acoustic parameters from a large German-speaking male sample. Explicit speech stereotypes for straight and gay German men complement each other. Voice pitch (“low” – “high”), nasality (“non-nasal” – “nasal”), chromaticity (“dark” – “bright”), and smoothness (“hard” – “soft”) were most important. Especially with regard to the belief that different pitches characterize gay and straight men’s voices, earlier findings for American English (Piccolo, 2008; Panfili, 2011) and Puerto Rican Spanish were replicated (Mack, 2010). In line with popular cultural images of gay men in Germany, we provided the first evidence that nasality is an equally important voice stereotype regarding men’s sexual orientation as voice pitch. Hereby, intercultural differences become apparent. In contrast to German speech stereotypes, previous American English studies emphasized the prominence of the “gay lisp” (see Munson & Babel, 2007) while nasality is disregarded as a relevant voice feature for characterizing gay and straight men’s speech in Singaporean English (Chiang, 2003). Hence, the “gay nasal” appears to be the German equivalent of the American English “gay lisp”.

Implicit speech stereotypes were investigated by regressing raters’ judgments of speakers’ sexual orientation on speakers’ acoustic parameters. While findings on the implicit level for median f0 supported the explicitly expressed voice pitch stereotype and evidence for CoG in /s/ and mean F2 corresponds to the chromaticity stereotype (associating lower values with straightness), the explicitly stated nasality difference for gay and straight men was not replicated on the implicit level. The evidence we found on the implicit level is fully in line with gender inversion theories. Compared to a previous investigation on acoustic correlates of perceived sexual orientation in male

German speakers, evidence on mean F2 was replicated whereas null findings for central tendency measures of f0 and /s/ were not (Sulpizio et al., 2015). Hence, mean F2 seems to be most important in driving the perception of men's sexual orientation.

Actual sexual orientation was signaled by only one acoustic parameter: Straight men showed lower mean F1 than gay men. This contradicts previous findings for mean F1 from studies using English speakers (Pierrehumbert et al., 2004, Munson et al., 2006; Rendall et al., 2008) but is compatible with former evidence for vowel-specific F1-values from a male German sample (Sulpizio et al., 2015; mean F1 was not analyzed in this paper). Thus, in contrast to German women (Kachel et al., 2017), there seems to be a relevant acoustic parameter that differentiates between male sexual orientation groups: F1. However, listeners neither use mean F1 for judging speaker's sexual orientation nor could it be linked to any explicitly expressed speech stereotypes. Because neither explicit nor implicit stereotypes correspond to actual differences, there is strong evidence for stereotypic inaccuracy as one aspect of speech stereotype inaccuracy.

Although there was an acoustic difference between actually gay and straight men, evidence for within-group differences was much stronger. This suggests a second type of stereotype inaccuracy: dispersion inaccuracy. The first piece of evidence showing acoustically potent variability within groups refers to the degree of sexual orientation. Corresponding to gender inversion theory, exclusively straight men showed lower center of gravity in /s/ than mainly straight men. However, contradicting gender inversion theory, exclusively gay men exhibited less jitter than mainly gay men. In line with a study on Dutch speakers (Baeck et al., 2011), we found no acoustically potent within-group differences for gay men on mean f0. Kachel et al. (2017) investigated the effects of actual sexual orientation on German women's speech and suggested that the acoustic potency of sexual orientation depends on contextual factors such as gender of the experimenter. Hence, we recommend future sociophonetic studies on men's speech to apply a fine-grained measurement of actual sexual orientation instead of mere di- or trichotomous sexual self-categorizations and examine the effects of experimenter's gender as well.

The second piece of evidence for acoustically potent variability within groups regards the question as to which psychological characteristics of gay men are associated with which (if any) acoustic parameters. In line with gender inversion theory, less contact to girls during childhood, more gender conforming voice self-assessment, and fewer discrimination experiences were connected to gender conforming acoustic parameters. This is in line with findings showing that coming-out age was not reflected in men's speech (Baeck et al., 2011). Furthermore, this evidence corresponds to findings for German lesbians that show their psychological characteristics being mirrored in their speaking patterns (Kachel et al., 2017). Hence, the present study emphasizes the relevance of including psychological characteristics when investigating men's speech, particularly to test multiple hypotheses on the origin and development of acoustic correlates of sexual orientation.

The third piece of evidence showing acoustically potent variability within groups refers to actual masculinity/femininity. We found gay and straight men who self-described as more masculine exhibiting lower mean F2. In contrast to previous studies (Biemans, 2000; Chiang, 2003), we relied on several speakers representing each group and used instruments that reliably and validly measure gender-role self-concept (for a recent discussion on the value and test quality criteria of gender-role self-concept instruments in general and the TMF in particular see Kachel et al., 2016). Hence, the present investigation provides the first evidence that actual masculinity/femininity is reflected in men's speech and is not a mere artifact of sexual orientation. Moreover, different acoustic parameters signal actual sexual orientation (mean F1) and masculinity/femininity (mean F2); however both are connected to vowel space. Whether mean F2 is used across different languages to indexicalize actual masculinity/femininity could be tested by future studies.

Overall, our approach for investigating men's speech has integrated two perspectives. When dealing with gay and straight men we investigated social categories ("linguistics of community") and we took a diversity-oriented focus when looking at intra-group variability and considering sexual orientation groups not only as externally definable categories ("linguistics of contact", see Barrett, 1997). We found evidence for acoustically potent variability within groups indicating that gay and straight men should be considered as internally heterogeneous, supporting conclusions drawn by Kachel et al. (2017) for lesbians and straight women. Hence, although lay gender inversion theory was partially supported, when looking in detail, the overall result pattern contradicts this theory because it assumes homogeneity within groups.

There are some limitations of the present study. Despite our relatively large group sizes ($n >$ 25 per group), we were only able to detect large effects. However, generalizations of findings regarding sexual orientation are limited, because nothing is known about how acoustic, psychological, and other characteristics are distributed in non-straight populations (Sandfort, 1997). Explicit calls for gay men's participation may provoke a self-selection of those feeling comfortable with their minority status and primed them to behave in a specific manner. Although we explicitly referred to sexual orientation during our recruitment, we did not make the sexual orientation of our speakers salient to them on the recording day. Letting our speakers fill out the Kinsey-like scale again before the recording, could have led to more pronounced speech differences due to activated group identity (see Tajfel & Turner, 1979; Hornsey, 2008). In order to increase statistical power, we recommend future studies to use even larger sample sizes and make speaker's group identity salient before the recording.

Additionally, our recordings were done in a lab by experimenters with prior knowledge of speakers' sexual orientation. Hence, generalization to everyday speaking behavior is problematic because of the artificiality of the recording environment and because the experimenter may have subconsciously treated gay and straight speakers differently. Future studies should reduce artifacts of lab speech by investigating natural surroundings by using mobile microphones (Podesva, 2011) and rely on experimenters blind to speakers' sexual orientation.

The present study has yielded two main findings. First, the implicit and explicit speech stereotypes our participants held about gay and straight men contained little truth because their stereotypic content was inaccurate (i.e., neither the explicit nor the implicit stereotype reflected the actual F1 difference found between gay and straight men). Second, perhaps even more importantly: Differences between gay and straight men are less important than differences *within* both groups, indicating dispersion inaccuracy. Hence, we found supporting scientific evidence for one of the main messages of the illustrative documentary movie “Do I sound gay?”: Speech stereotypes regarding male sexual orientation seem to be inaccurate.

Acknowledgments

Substantial contributions to the conception or design of the work: SK, APS, MCS; substantial contribution to the acquisition and analysis of the data: SK, APS, MCS; interpretation of data for the work: SK, APS, MCS; Drafting the work or revising it critically for important intellectual content: SK, APS, MCS; Final approval of the version to be published: SK, APS, MCS.

The current research was funded by a grant from the German Research Foundation to the last two authors (DFG STE 938/10-2; FOR 1097). We thank Monique Beck, Maren Schlehe, Rebekka Aust, Alexander Makosch, Felix Sokoll, and Felicia Schuld for help with data collection and data pre-processing. The authors declare no conflicts of financial, personal, or other interests.

References

Allport, G. W. (1954). *The nature of prejudice*. Massachusetts: Addison-Wesley.

Baeck, H., Corthals, P., & van Borsel, J. (2011). Pitch characteristics of homosexual males. *Journal of Voice*, 25, 211-214.

Barrett, R. (1997). The 'homo-genius' speech community. In A. Livia & K. Hall (Eds.), *Queerly phrased. Language, gender, and sexuality,* (pp. 181-201). New York: Oxford University Press.

Biemans, M. A. J. (2000). *Gender variation in voice quality* (Doctoral dissertation, Catholic University Nijmegen, Nijmegen. Netherlands). Retrieved from http://www.lotpublications.nl/Documents/38_fulltext.pdf

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, 5, 341-345.

Bruckmüller, S. (2013). Singled Out as the Effect to be Explained: Implications for Collective Self-Esteem. *Personality and Social Psychology Bulletin,* 39, 237–249. DOI: 10.1177/0146167212471686.

Camp, M. (2009). *Japanese lesbian speech: Sexuality, gender identity, and language* (Doctoral dissertation, University of Arizona). Retrieved from http://hdl.handle.net/10150/195371

Chiang, S. W. W. (2003). *Sounding 'gay': Prosodic patterns in gay speech in Singapore.* (Master thesis, National University of Singapore, Singapore) Retrieved from http://scholarbank.nus.edu.sg/handle/10635/14832

Fletcher, S. G. (1970). Theory and instrumentation for quantitative measurement of nasality. *Cleft Palate Journal,* 7, 601-609.

Gaudio, R. P. (1994). Sounding Gay: Pitch Properties in the Speech of Gay and Straight Men. *American Speech,* 69, 30-57.

Gertler, H., Thorpe , D. (Producers), & Thorpe, D. (Director) (2014). *Do I sound gay?* [Motion picture]. United States of America.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review,* 102, 4-27. doi:10.1037/0033-295X.102.1.4

Hancock, A., Colton, L., & Douglas, F. (2014). Intonation and gender perception: Applications for transgender speakers. *Journal of Voice,* 28, 203–209.

Herek, G. M., & Berrill, K. T. (1990). Documenting the victimization of lesbians and gay men: Methodological issues. *Journal of Interpersonal Violence*, 5, 301-315.

Hodson, G., Harry, H., & Mitchell, A. (2009). Independent benefits of contact and friendship on attitudes toward homosexuals among authoritarians and highly identified heterosexuals. *European Journal of Social Psychology*, 39, 509-525.

Hornsey, M. J. (2008). Social identity theory and self-categorization theory: A historical review. *Social and Personality Psychology Compass*, 2, 204–222. DOI: 10.1111/j.1751-9004.2007.00066.x

Jacobs, G. (1996). Lesbian and gay male language use: A critical review of the literature. *American Speech*, 71, 49-71.

Judd, C. M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review,* 100, 109–128. DOI: 10.1037/0033-295X.100.1.109

Jussim, L., Crawford, J. T., & Rubinstein, R. S. (2015). Stereotype (in)accuracy in perceptions of groups and individuals. *Current Directions in Psychological Science*, 24, 490–497.

Kachel, S., Simpson, A. P., & Steffens, M. C. (2017). Acoustic correlates of sexual orientation and gender-role self-concept in women's speech. re-submitted to Journal of the Acoustical Society of America.

Kachel, S., Steffens, M. C., & Niedlich, C. (2016). Traditional masculinity and femininity: Validation of a new scale assessing gender roles. *Frontiers in Psychology*, 7, Article 956. doi: 10.3389/fpsyg.2016.00956.

Kinsey, A., Pomeroy, W., & Martin, C. (1948). *Sexual Behavior in the Human Male* (Saunders, Philadelphia), 1-804.

Kite, M., & Deaux, K. (1987). Gender belief systems: Homosexuality and the implicit inversion theory. *Psychology of Women Quarterly*, 11, 83-96.

Kohler, K. J. (1994). Glottal stops and glottalization in German. Data and theory of connected speech processes. *Phonetica*, 51, 38-51.

Lerman, J., & Damsté, P. (1969). Voice pitch of homosexuals. *Folia Phoniatrica et Logopaedica,* 21, 340-346.

Li, F., Rendall, D., Vasey, P. L., Kinsman, M., Ward-Sutherland, A., & Diano, G. (2016). The development of sex/gender-specific /s/ and its relationship to gender identity in children and adolescents. *Journal of Phonetics*, 57, 59-70.

Linville, S. E. (1998). Acoustic Correlates of Perceived versus Actual Sexual Orientation in Men's Speech. *Folia Phoniatrica et Logopaedica,* 50, 35-48.

Lippa, R. A. (2005). Sexual orientation and personality. *Annual Review of Sex Research*, 16, 119-153.

Luhtanen, R., & Crocker, J. (1992). A Collective Self-Esteem Scale: Self-evaluation of one's social identity. *Personality and Social Psychology Bulletin*, 18, 302-318.

Mack, S. (2010). Perception and Identity: Stereotypes of Speech and Sexual Orientation in Puerto Rican Spanish. In C. Borgonovo, M. Español-Echevarría, & P. Prévost (Eds.), Selected Proceedings of the 12th Hispanic Linguistics Symposium (136-147). Cascadilla Proceedings Project. http://www.lingref.com/cpp/hls/12/paper2412.pdf

Madon, S. (1997). What do people believe about gay males? A study of stereotype content and strength. *Sex Roles,* 37, 663–685.

Mohr, J. J., & Kendra, M. S. (2011). Revision and extension of a multidimensional measure of sexual minority identity: The Lesbian, Gay, and Bisexual Identity Scale. *Journal of Counseling Psychology,* 58, 234-245.

Munson, B., & Babel, M. (2007). Loose Lips and Silver Tongues, or, Projecting Sexual Orientation Through Speech. *Language and Linguistics Compass,* 1, 416-449. DOI: 10.1111/j.1749-818x.2007.00028.x.

Munson, B., McDonald, E. C., DeBoe, N. L., & White, A. R. (2006). The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. Journal of Phonetics, 34, 202-240.

Munson, B. (2007). The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation. *Language and Speech,* 50, 125–142. DOI: 10.1177/00238309070500010601.

Panfili, L. (2011). Vowel Duration and Perceptions of the Gay Accent. *Linguistics Honors Projects,* Paper 7. Retrieved from http://digitalcommons.macalester.edu/ling_honors/7/

Piccolo, F. (2008). Listener-identified phonetic correlates of gay-, lesbian- and straight-sounding speech (Doctoral dissertation, University of Hawaii). Retrieved from https://scholarspace.manoa.hawaii.edu/handle/10125/20680

Pierrehumbert, J. B., Bent, T., Munson, B., Bradlow, A. R., & Bailey, M. J. (2004). The influence of sexual orientation on vowel production (L). *Journal of the Acoustical Society of America*, 116, 1905-1908.

Podesva, R. J. (2011). The California vowel shift and gay identity. *American Speech,* 86, 32-51.

Rendall, D., Vasey, P. L., & McKenzie, J. (2008). The queen's English: An alternative, biosocial hypothesis for the distinctive features of 'gay speech'. *Archives of Sexual Behavior*, 37, 188-204.

Rieger, G., Linsenmeier, J. A. W., Gygax, L., & Bailey, J. M. (2008). Sexual orientation and childhood gender nonconformity: Evidence from home videos. Developmental Psychology, 44, 46-58.

Rogers, H., & Smyth, R. (2003). Phonetic Differences Between Gay- and Straight-Sounding Male Speakers of North American English. *15th International Congress of Phonetic Sciences*, 1855-1858. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/p15_1855.html

Runge, T. E., Frey, D., Gollwitzer, P. E., Helmreich, R. L., & Spence, J. T. (1981). Masculine (instrumental) and feminine (expressive) traits: A comparison between students in the United

States and West Germany. *Journal of Cross-Cultural Psychology*, 12, 142-162.

Sandfort, T. G. M. (1997). Sampling male homosexuality. In J. Bancroft (Ed.), *Researching sexual behavior: Methodological issues,* (pp. 261-275). Bloomington: Indiana University Press.

Schubert, T. W., & Otten, S. (2002). Overlap of self, ingroup, and outgroup: Pictorial measures of self-categorization. *Self and Identity*, 1, 353-376.

Shively, M. G., Jones, C., & DeCecco, J. P. (1985). Research on sexual orientation: Definitions and methods. *Journal of Homosexuality*, 9, 127-136

Simpson, A. P., Kohler, K. J., & Rettstadt, T. (1997). The Kiel Corpus of Read/Spontaneous Speech — Acoustic data base, processing tools and analysis results. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK),* 32.

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass,* 3, 621-640. doi: 10.1111/j.1749-818x.2009.00125.x.

Sisson, C. (2003). Phonetic Cues in the Evaluation of Gay Male Speech in Canadian English and Québec French. Retrieved from http://www.cs.cmu.edu/~./csisso/docs/SocioII-paper.pdf

Smyth, R., Jacobs, G., & Rogers, H. (2003). Male voices and perceived sexual orientation: An experimental and theoretical approach. *Language in Society*, 32, 329-350.

Smyth, R., & Rogers, H. (2008). Do gay-sounding men speak like women? *Toronto Working Papers in Linguistics*, 27, 129-144.

Spence, J. T., & Helmreich, R. L. (1978). *Masculinity and femininity: Their psychological dimensions, correlates and antecedents,* Austin: University of Texas Press.

Steffens, M. C., Bergert, M., & Heinecke, S. (2010). *Doppelt diskriminiert oder gut integriert? Zur Lebenssituation von Lesben und Schwulen mit Migrationshintergrund (Discriminated twice or well integrated? The living situation of lesbians and gay men with a migration background).* Berlin: LSVD. Retrieved from http://www.migrationsfamilien.de/pdf/Studie_Jena.pdf

Stuart-Smith, J. (2007). Empirical evidence for gendered speech production: /s/ in Glaswegian. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology 9*, (pp. 65–86). New York: Mouton de

Gruyter.

Sulpizio, S., Fasoli, F., Maass, A., Paladino, M. P., Vespignani, F., Eyssel, F., & Bentler, D. (2015). The Sound of Voice: Voice-Based Categorization of Speakers' Sexual Orientation within and across Languages. *PLoS ONE*, 10, e0128882. doi: 10.1371/journal.pone.0128882

Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *Psychology of Intergroup Relations* (pp. 33–47). Monterey, Calif: Brooks/Cole Pub. Co.

Valentova, J. V., & Havlíček, J. (2013). Perceived Sexual Orientation Based on Vocal and Facial Stimuli Is Linked to Self-Rated Sexual Orientation in Czech Men. *PLoS ONE,* 8, e82417. doi: 10.1371/journal.pone.0082417.

Weirich, M., & Simpson, A. P. (2013). Investigating the relationship between average speaker fundamental frequency and acoustic vowel space size. *Journal of the Acoustical Society of America*, 134, 2965-2974.

Williams, D. R., Yu, Y., Jackson, J. S., & Anderson, N. B (1997). Racial differences in physical and mental health: Socioeconomic status, stress, and discrimination. *Journal of Health Psychology*, 2, 335-351.

Worthington, R. L., & Reynolds, A. L. (2009). Within-group differences in sexual orientation and identity. *Journal of Counseling Psychology*, 56, 44-55.

Zimman, L. (2010). Female-to-male transsexuals and gay-sounding voices: A pilot study. *Colorado Research in Linguistics*, 22.

Zwicky, A. M. (1997). Two lavender issues for linguists. In A. Livia & K. Hall (Eds.), *Queerly phrased. Language, gender, and sexuality,* (pp. 21-34). New York: Oxford University Press.

Tables

Table 1. Overview of Previous Studies on Explicit Speech Stereotypes and Acoustic Correlates of Men's Perceived and Actual Sexual Orientation.

	Explicit speech stereotypes		Implicit speech stereotypes (listener-centered studies on sexual orientation)		Kernel of truth (speaker-centered studies on sexual orientation)	
Acoustic parameter	Deductions from GID[3]	Empirical findings	Supporting evidence	No evidence	Supporting evidence	No evidence
Intonational features						
F0 measure of central tendency[1]	SM < GM	SM: low pitch; GM: high pitch (Piccolo, 2008; Mack, 2010; Panfili, 2011)	Linville (1998); Sisson (2003): CEM; Valentova and Havlíček (2013): rated by gay men	Gaudio (1994); Rogers and Smyth (2003); Sisson (2003): CFM; Smyth et al. (2003); Munson et al. (2006); Valentova and Havlíček (2013): rated by straight women; Sulpizio et al. (2015)	Linville (1998); Baeck et al. (2011)	Lerman and Damsté (1969); Gaudio (1994); Sisson (2003); Munson et al. (2006); Rendall et al. (2008); Zimman (2010); Valentova and Havlíček (2013); Sulpizio et al. (2015)
F0 range	SM < GM	SM: flat tone; GM: broad pitch range (Piccolo, 2008)	Sisson (2003): CEM	Gaudio (1994); Rogers and Smyth (2003); Sisson (2003): CFM; Munson et al. (2006)	-	Gaudio (1994); Munson et al. (2006); Zimman (2010)
Sibilant measures						
Measures of central tendency for /s/[2]	SM < GM	GM: lisping (Piccolo, 2008; Mack, 2011; Panfili, 2011)	Linville (1998); Sulpizio et al. (2015): IM	Munson et al. (2006); Zimman (2010); Sulpizio et al. (2015): GM	Linville (1998)	Zimman (2010)
/s/ skewness	SM > GM		Munson et al. (2006)	Sulpizio et al. (2015)	Munson et al. (2006)	Sulpizio et al. (2015)
Measures of central tendency for /ʃ/[2]	SM < GM	-	-	Munson et al. (2006); Zimman (2010)	-	Munson et al. (2006); Zimman (2010)
/ʃ/ skewness	SM > GM	-	-	Munson et al. (2006)		Munson et al. (2006)
Vowel space characteristics						

Vowel space shift (mean F1 and F2)	SM < GM	-	Munson et al. (2006)	Piccolo (2008)[4]; Zimman (2010)	-	Pierrehumbert et al. (2004); Munson et al. (2006); Rendall et al. (2008); Zimman (2010)
Vowel space expansion	SM < GM	SM: poor articulation; GM: precise pronunciation (Piccolo, 2008; Mack, 2011; Panfili, 2011)	-	Munson et al. (2006); Piccolo (2008)	Pierrehumbert et al. (2004)	Munson et al. (2006); Rendall et al. (2008)
F1 and F2 in single vowels	SM < GM	-	-	Sulpizio et al. (2015)	Pierrehumbert et al. (2004); Munson et al. (2006); Rendall et al. (2008) Sulpizio et al. (2015)	Pierrehumbert et al. (2004)[4]; Munson et al. (2006); Rendall et al. (2008)[4]; Zimman (2010); Sulpizio et al. (2015)
Temporal characteristics						
Segmental durational differences	SM < GM	GM: longer s (Panfili, 2011)	Linville (1998); Sisson (2003): CFM; Munson et al. (2006); Sulpizio et al. (2015)	Sisson (2003): CEM; Piccolo (2008); Zimman, 2010; Panfili (2011)	Linville (1998)	Sisson (2003); Pierrehumbert et al. (2004); Munson et al. (2006); Zimman (2010); Sulpizio et al. (2015)
Voice onset time	SM < GM	-				
Voice quality parameters						
Breathiness, Creakiness	SM: creaky; GM: breathy	GM: soft, breathy (Piccolo, 2008)	-	Munson et al. (2006)	-	Munson et al. (2006)
Nasalance	-	GM: nasal voice (Mack, 2010)				

Abbreviation: Canadian English men (CEM), Canadian French men (CFM), gender inversion theory (GID), Italian men (IM), German men (GM), Gay men (GM), Straight men (SM).
[1]f0 mean and modal speaking frequency. [2]peak frequency and center of gravity. [3]As a basis for most comparisons served a review on gender divergent speaking patterns (Simpson, 2009). [4]Contradicting evidence instead of or in addition to non-significant finding.

Table 2. Psychological Differences between Gay and Straight Speakers (comparison 1), between Exclusively and Mainly Gay (comparison 2), and between Mainly and Exclusively Straight Speakers (comparison 3).

	Comparison 1			Comparison 2			Comparison 3		
Psychological characteristics	Gay[a]	Straight[b]	p-value	Exclusiv. gay[c]	Mainly gay[d]	p-value	Mainly straight[f]	Exclusiv. straight[g]	p-value
Sexual Orientation									
[h]Gyno-Androphily-Difference	**4.79**	**4.72**	**<.001**	-4.94	-4.47	.246	4.07	4.96	.042
[i]Perceived Straightness	.57	.71	.028	.60	.51	.354	.69	.71	.773
Gender-role self-concept									
[j]TMF	**3.49**	**2.51**	**<.001**	3.43	3.60	.651	2.90	2.37	.225
[k]GEPAQ-F	4.05	3.72	.027	4.03	4.07	.821	3.92	3.65	.246
[k]GEPAQ-M	3.31	3.41	.601	3.37	3.20	.573	3.47	3.38	.733
[k]CGNS	**3.03**	**1.78**	**<.001**	2.95	3.22	.398	1.97	1.71	.226
Group affiliation to…									
[j]…gay men	**4.64**	**2.42**	**<.001**	4.59	4.75	.743	3.43	2.05	.004
[j]…straight men	**3.44**	**5.62**	**<.001**	3.41	3.50	.874	5.00	5.84	.315
[j]**Psycho-social identification**	4.38	4.80	.185	4.24	4.67	.385	4.06	5.08	.037
Social environment									
[j]CurCon to gay men	**5.16**	**3.85**	**.001**	5.18	5.13	.883	3.86	3.84	.983
[j]CurCon to straight men	**5.48**	**6.48**	**.001**	5.24	6.00	.150	6.67	6.42	.434
[j]GenDis of friends	**3.64**	**4.65**	**<.001**	3.53	3.88	.221	4.43	4.74	.504
[j]SexOr of male friends	**4.00**	**6.23**	**<.001**	4.00	4.00	1.000	5.71	6.42	.004
[j]ConGirls	**5.76**	**4.50**	**.001**	5.94	5.38	.198	4.71	4.42	.679
[j]ConBoys	**4.92**	**6.35**	**<.001**	5.12	4.50	.260	5.86	6.53	.056
Voice self-assessment	**4.28**	**5.44**	**<.001**						
Coming-out related characteristics									
[j]Coming-Out extent				5.42	4.89	.322			
[k]Concealment				2.21	2.38	.634			

motivation			
[l]Discrimination	1.44	2.06	.071

Abbreviations: Traditional Masculinity-Femininity Scale (TMF), German Extended Personality Attributes Questionnaire (GEPAQ), Gender Role Behavior Scale (GRB) and Childhood Gender Nonconformity Scale (CGNS). Endings indicate masculinity (-M) and femininity (-F) scales. Contact to corresponding gender during childhood (ChildCon), current contact (CurCon), sexual orientation (SexOr), and gender distribution (GenDis).

[a] n = 25 with Kinsey-like scores for gay men 1-2. [b] n = 26 with Kinsey-like scores for straight men 6-7. [c] n = 17. [d] n = 8. [e] n = 7. [f] n = 19. [h] Scale ranged from -6 to 6. [i] Scale ranged from .00 to 1.00. [j] Scale ranged from 1 to 7. [k] Scale ranged from 1 to 6. [l] Scale ranged from 1 to 5.

The significance level is α = .0031 after adjusting it to 16 tests according to Bonferroni correction formula. Significant p-values are **bold**.

Table 3. Acoustic differences between gay and straight speakers (comparison 1), between exclusively and mainly gay (comparison 2), and between mainly and exclusively straight speakers (comparison 3), as well as acoustic correlates of present gender-role self-concept and perceived straightness.

	Comparison 1			Comparison 2			Comparison 3			TMF		Perceived Straightness	
Acoustic parameter	Gay[a]	Straight[b]	p-value[c]	Exclusiv. gay[d]	Mainly gay[e]	p-value[f]	Mainly straight[g]	Exclusiv. straight[h]	p-value[c]	*r*[i]	p-value[c]	*r*[i]	p-value[c]
Intonational characteristics (Hz)													
F0 median read	119	116	.467	118	122	.478	119	115	.509	-.07	.612	-.39	.005
F0 range read	89	81	.149	89	90	.937	87	79	.341	-.08	.579	-.21	.138
F0 median spon	116	110	.092	116	118	.748	115	109	.281	-.01	.954	**-.48**	**<.001**
F0 range spon	98	95	.727	99	95	.721	101	93	.445	-.10	.465	-.124	.376
Sibilant measures (Hz)													
CoG ʃ	3406	3312	.402	3400	3420	.920	3513	3238	.080	.30	.036	-.23	.109
Skewness ʃ	1.29	1.02	.052	1.25	1.39	.521	1.22	.96	.201	.32	.022	-.31	.025
CoG s	6308	5699	.010	6302	6321	.952	6256	5493	[*].045	**.46**	**.001**	**-.47**	**<.001**
Skewness s	-0.51	-0.02	.009	-.46	-.61	.506	-.08	.00	.749	-.37	.008	.36	.010
CoG-Diff s and ʃ	2902	2387	.012	2902	2902	1.000	2743	2255	.131	.37	.007	-.43	.002
Vowel space features (Hz)													
F1 in /iː/	273	256	.016	272	274	.895	256	257	.887	.29	.038	-.21	.140
F2 in /iː/	2182	2110	.039	2188	2168	.696	2110	2110	.991	.29	.037	-.29	.042

F1 in /a:/	**696**	**653**	**.001**	696	695	.939	668	647	.317	.24	.087	-.38	.006
F2 in /a:/	1260	1200	.014	1266	1249	.582	1228	1189	.383	.41	.003	**-.44**	**.001**
F1 in /u:/	289	277	.037	291	286	.549	274	279	.613	.29	.042	-.29	.041
F2 in /u:/	661	661	.998	650	685	.167	674	656	.739	.37	.008	-.23	.099
Mean F1	**419**	**395**	**<.001**	420	418	.856	399	394	.531	.37	.008	**-.44**	**.001**
Mean F2	1368	1324	.020	1368	1367	.978	1337	1320	.586	**.49**	**<.001**	**-.45**	**.001**
VowelSpace	.32	.28	.023	.19	.31	.600	.29	.27	.514	.12	.395	-.25	.080
Plosive characteristics (ms)													
VOT /p/	45.44	42.73	.516	44.94	47.50	.830	45.29	41.79	.555	.19	.176	-.14	.328
Asp /p/	61.52	58.81	.451	62.12	60.25	.755	58.57	58.89	.953	.36	.009	-.12	.405
VOT /t/	68.16	64.27	.405	67.76	69.00	.870	62.71	64.84	.771	.31	.026	-.35	.011
Asp /t/	83.68	78.42	.236	84.82	81.25	.579	70.43	81.37	.120	.15	.281	-.20	.151
VOT /k/	66.76	62.77	.385	66.41	67.50	.894	60.43	63.63	.611	.26	.062	.15	.014
Asp /k/	81.24	79.53	.700	81.82	80.00	.801	75.57	81.00	.425	.26	.071	-.36	.009
Voice quality parameters													
Jitter /a:/ in %	.29	.47	.009	.26	.34	*.037	.69	.39	.164	-.12	.375	.09	.087
Shimmer /a:/ in %	1.73	2.11	.154	1.65	1.89	.295	3.18	1.72	.085	-.16	.264	.00	.991
HNR /a:/ in dB	21.59	19.64	.019	22.00	20.70	.124	17.62	20.38	.176	.17	.211	-.11	.449
Nasalance in %	33.92	35.23	.633							-.16	.324	.12	.465

Abbreviations: Center of gravity (CoG), read speech in butter story (read), spontaneous speech during path description (spon), voice onset time (VOT), aspiration (Asp), harmonics-to-noise ratio (HNR).

[a] $n = 25$ with Kinsey-like scores for gay men 1-2. [b] $n = 26$ with Kinsey-like scores for straight men 6-7. [c] The significance level is $\alpha = .0018$ after adjusting it to 28 tests according to the Bonferroni correction formula. Significant p-values are **bold**. [d] $n = 17$. [e] $n = 8$. [f] The significance level was left unadjusted for exploratory reasons. Significant p-values were indicated by the common levels of asterisks. [g] $n = 7$. [h] $n = 19$. [i] $n = 51$.

$^{*}p \leq .05$. $^{**}p \leq .01$. $^{***}p \leq .001$.

Table 4. Regression of Perceived Sexual Orientation on Acoustic Parameters Showing Significant Correlations with Perceived Sexual Orientation.

Predictors and model fit criteria	Step 1				Step 2				Step 3			
	b	*SE*	Beta	*p*	*b*	*SE*	Beta	*p*	*b*	*SE*	Beta	*p*
Predictors												
F0 median in spon	-.01	.00	-.48	<.001	-.01	.00	-.40	.001	-.01	.00	-.35	.003
CoG /s/					.00	.00	-.40	.001	.00	.00	-.33	.005
Mean F2									.00	.00	-.29	.013
Model fit												
Corrected R^2	$R^2 = .21$				$R^2 = .35$				$R^2 = .42$			
Significance of R^2-change	$F(1, 48) = 14.22$, $p < .001$				$F(1, 47) = 11.50$, $p = .001$				$F(1, 46) = 6.72$, $p < .013$			

Abbreviations: Center of gravity (CoG), spontaneous speech during path description (spon).

Excluded variables: F2 in /a:/ and mean F1 (both Beta ≤ -.12, both ps ≥ -.58)

*Table 5. Correlations of Psychological Characteristics and Acoustic Parameters within the Gay Group (*n *= 25).*

Acoustic parameters	GEPAQ-F+	CGNS	ConGirls	Voice self-assessment	Discrimination
Intonational characteristics (Hz)					
F0 median read	-.24 (.250)	.16 (.441)	.18 (.395)	-.42 (.074)	.26 (.215)
F0 range read	-.31 (.129)	-.00 (.940)	.31 (.133)	-.23 (.347)	-.04 (.839)
F0 median spon	-.17 (.426)	.29 (.172)	.36 (.082)	*-.56 (.015)	.15 (.476)
F0 range spon	-.23 (.279)	-.00 (.991)	*.50 (.013)	-.43 (.078)	-.02 (.938)
Sibilant measures (Hz)					
CoG ʃ	*.46 (.022)	.21 (.319)	.23 (.275)	-.20 (.407)	-.03 (.889)
Skewness ʃ	.26 (.204)	.27 (.200)	-.17 (.406)	-.24 (.327)	.27 (.187)
CoG s	*.42 (.037)	*.40 (.048)	.09 (.684)	-.35 (.137)	.06 (.790)
Skewness s	-.07 (.750)	-.18 (.401)	-.23 (.270)	.35 (.145)	-.12 (.567)
CoG-Diff s and ʃ	.16 (.435)	.30 (.145)	-.05 (.805)	-.22 (.370)	.08 (.706)
Vowel space features (Hz)					
F1 in /iː/	.05 (.826)	*.40 (.046)	.15 (.466)	-.26 (.284)	*.42 (.037)
F2 in /iː/	.07 (.741)	.36 (.076)	.14 (.514)	.03 (.896)	.27 (.198)
F1 in /aː/	-.09 (.674)	.15 (.463)	.31 (.139)	-.44 (.062)	.05 (.803)
F2 in /aː/	.04 (.865)	-.02 (.916)	-.27 (.201)	-.22 (.375)	.22 (.297)
F1 in /uː/	*.45 (.023)	.23 (.267)	.18 (.381)	-.29 (.234)	.05 (.814)
F2 in /uː/	*.42 (.037)	.10 (.625)	.13 (.524)	-.35 (.145)	*.45 (.026)
Mean F1	.10 (.621)	.37 (.071)	.34 (.095)	*-.53 (.019)	.25 (.234)
Mean F2	.22 (.298)	.29 (.165)	.03 (.876)	-.23 (.350)	*.45 (.026)
VowelSpace	-.20 (.345)	.13 (.549)	.16 (.451)	-.10 (.700)	-.03 (.876)
Plosive characteristics (ms)					

VOT /p/	.28 (.173)	.12 (.575)	-.18 (.384)	-.19 (.428)	.05 (.830)
Asp /p/	.38 (.065)	.14 (.507)	-.14 (.505)	-.33 (.163)	.14 (.513)
VOT /t/	**.51 (.009)	*.43 (.030)	-.07 (.757)	-.09 (.723)	.04 (.850)
Asp /t/	.20 (.338)	.15 (.479)	.18 (.387)	-.24 (.318)	.03 (.893)
VOT /k/	.34 (.099)	.10 (.652)	-.28 (.182)	.14 (.572)	-.05 (.799)
Asp /k/	*.40 (.048)	.08 (.947)	-.15 (.474)	-.03 (.895)	-.02 (.934)
Voice quality parameters					
Jitter /a:/ in %	.01 (.949)	-.04 (.863)	-.14 (.447)	-.33 (.170)	.39 (.054)
Shimmer /a:/ in %	-.30 (.151)	-.37 (.064)	*-.42 (.036)	-.03 (.903)	-.06 (.788)
HNR /a:/ in dB	.09 (.668)	.06 (.772)	.26 (.211)	*.46 (.046)	-.09 (.668)
Nasalance in %	.12 (.631)	.10 (.704)	**-.59 (.010)	.05 (.857)	.36 (.140)

Abbreviations: Femininity scale of the German Extended Personality Attributes Questionnaire (GEPAQ-F), Childhood Gender Nonconformity Scale (CGNS), contact to girls durings childhood (ConGirls), center of gravity (CoG), read speech in butter story (read), spontaneous speech during path description (spon), voice onset time (VOT), aspiration (Asp), harmonics-to-noise ratio (HNR).

The significance level was left unadjusted due to exploratory reasons. Significant p-values are indicated by: $^{*}p \leq .05$. $^{**}p \leq .01$. $^{***}p \leq .001$.

Figures

Figure 1. Correlation of present gender-role self-concept (TMF) and mean F2 (upper panel) and center of gravity in /s/ (lower panel) separated for gay and straight men.

Figure 2. Explicit speech stereotypes regarding sexual orientation, arranged according to mean differences (for bipolar scales) and differences from scale midpoint (for unipolar scales).

Manuscript 3 – Related Appendix

Supplemental Material

Table A1. Distribution of Speakers Regarding Sexual Orientation on Kinsey-like Scale in Sessions 1 and 2.

	Number per cell	
Kinsey-like score	Session 1	Session 2
1 - Exclusively gay	17	15
2 - Mainly gay	8	4
3 - Somewhat gay	2	0
4 - Equally gay and straight	1	2
5 - Somewhat straight	0	0
6 - Mainly straight	7	5
7 - Exclusively straight	19	13

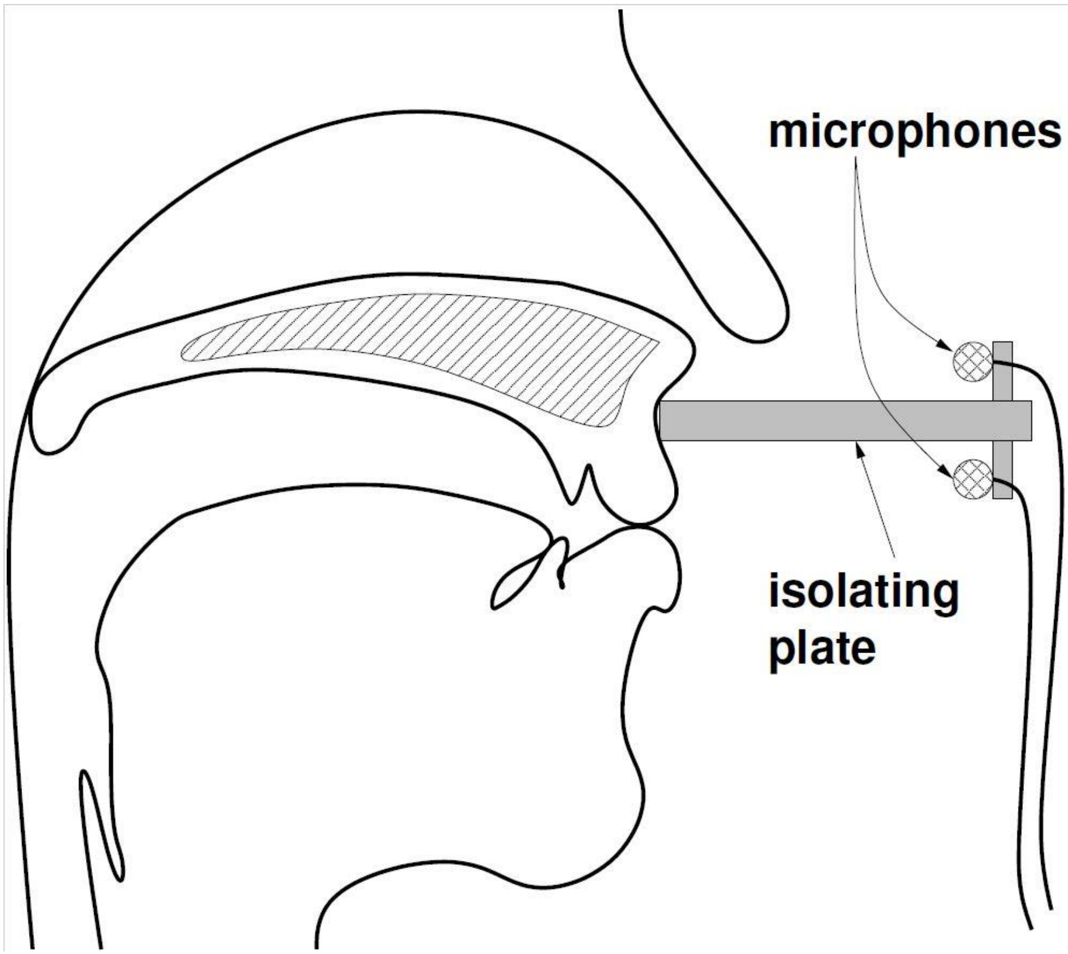


Figure A1. *Schematic sagittal view of the isolation plate and nasal and oral microphones.*

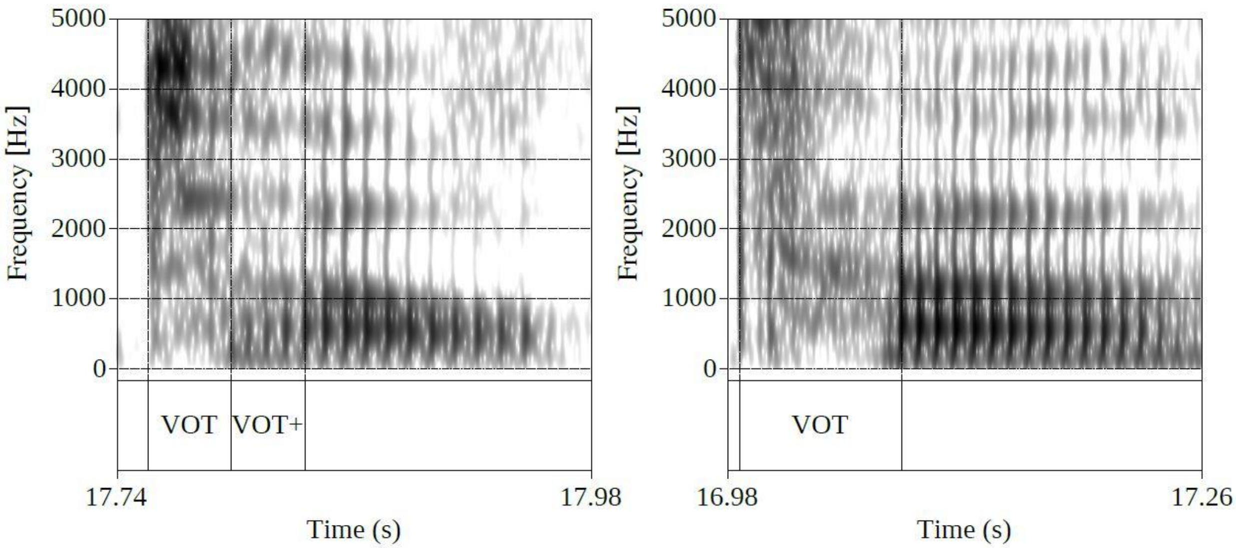


Figure A2. Spectrograms of the second syllable from two tokens of getaucht ("dived") with (left) and without (right) a stretch of breathy voice following VOT. The breathy voiced stretch is annotated with VOT+.

Online Questionnaire

Sexual orientation

Sexual self-identification[+] (Kinsey-like Scale; Kinsey, Pomeroy, & Martin, 1948).

[+]We adapted the original Kinsey-scale for our study with regard to the scale range and the description for every single scale point.

Item: "Regarding my sexual orientation I identify as...".

1	2	3	4	5	6	7
Exclusively gay	Mainly gay	Somewhat gay	Equally gay and straight	Somewhat straight	Mainly straight	Exclusively straight

S*exual orientation towards men and women (more objectifiable measure).*

"How often have you experienced…?"

	Never	Very seldom	Seldom	From time to time	Often	Very often	Always
	1	*2*	*3*	*4*	*5*	*6*	*7*
…sexual fantasies with men	○	○	○	○	○	○	○
…sexual fantasies with women	○	○	○	○	○	○	○
…romantic attraction towards men	○	○	○	○	○	○	○
…romantic attraction towards women	○	○	○	○	○	○	○
…physical attraction towards men	○	○	○	○	○	○	○
…physical attraction towards women	○	○	○	○	○	○	○
…sexual interaction with men	○	○	○	○	○	○	○
…sexual interaction with women	○	○	○	○	○	○	○

Masculinity/femininity

Traditional Masculinity-Femininity Scale (TMF by Kachel, Steffens, and Niedlich, 2016).

	Very masculine						Very feminine
	1	*2*	*3*	*4*	*5*	*6*	*7*
I consider myself…	○	○	○	○	○	○	○
Ideally, I would like to be…	○	○	○	○	○	○	○
Traditionally, my interests would be regarded as…	○	○	○	○	○	○	○
Traditionally, my attitudes and beliefs would be regarded as…	○	○	○	○	○	○	○
Traditionally, my behavior would be regarded as…	○	○	○	○	○	○	○
Traditionally, my outer appearance would be regarded as…	○	○	○	○	○	○	○

German Extended Personal Attributes Questionnaire Masculinity-Scale (GEPAQ-M by Runge, Frey, Gollwitzer, Helmreich, and Spence, 1981).

Not independent	*1*	*2*	*3*	*4*	*5*	Very independent
Very passive	*1*	*2*	*3*	*4*	*5*	Very active
Not competitive	*1*	*2*	*3*	*4*	*5*	Very competitive
*Decisive	*1*	*2*	*3*	*4*	*5*	Not decisive
Gives up easily	*1*	*2*	*3*	*4*	*5*	Never gives up
Not self-confident	*1*	*2*	*3*	*4*	*5*	Self-confident
Feels inferior	*1*	*2*	*3*	*4*	*5*	Feels superior
Doesn't stand up under pressure	*1*	*2*	*3*	*4*	*5*	Stands up under pressure

*Items with asterisks had to be recoded (1 to 5, 2 to 4, 3 to 3, 4 to 2, 5 to 1).

German Extended Personal Attributes Questionnaire Femininity-Scale (GEPAQ-F by Runge, Frey, Gollwitzer, Helmreich, and Spence, 1981).

Not emotional	*1*	*2*	*3*	*4*	*5*	Very emotional
*Devotes self to others	*1*	*2*	*3*	*4*	*5*	Doesn't devote self to others
Very rough	*1*	*2*	*3*	*4*	*5*	Very gentle
Not helpful	*1*	*2*	*3*	*4*	*5*	Very helpful
Very unkind	*1*	*2*	*3*	*4*	*5*	Very kind
Not aware of feelings	*1*	*2*	*3*	*4*	*5*	Aware of feelings
Not understanding	*1*	*2*	*3*	*4*	*5*	Very understanding
Cold	*1*	*2*	*3*	*4*	*5*	Warm

*Items with asterisks had to be recoded (1 to 5, 2 to 4, 3 to 3, 4 to 2, 5 to 1).

Gender-role self-concept during childhood

Childhood Gender Nonconformity Scale+ (CGNS, see Rieger, Linsenmeier, Gygax, & Bailey, 2008).

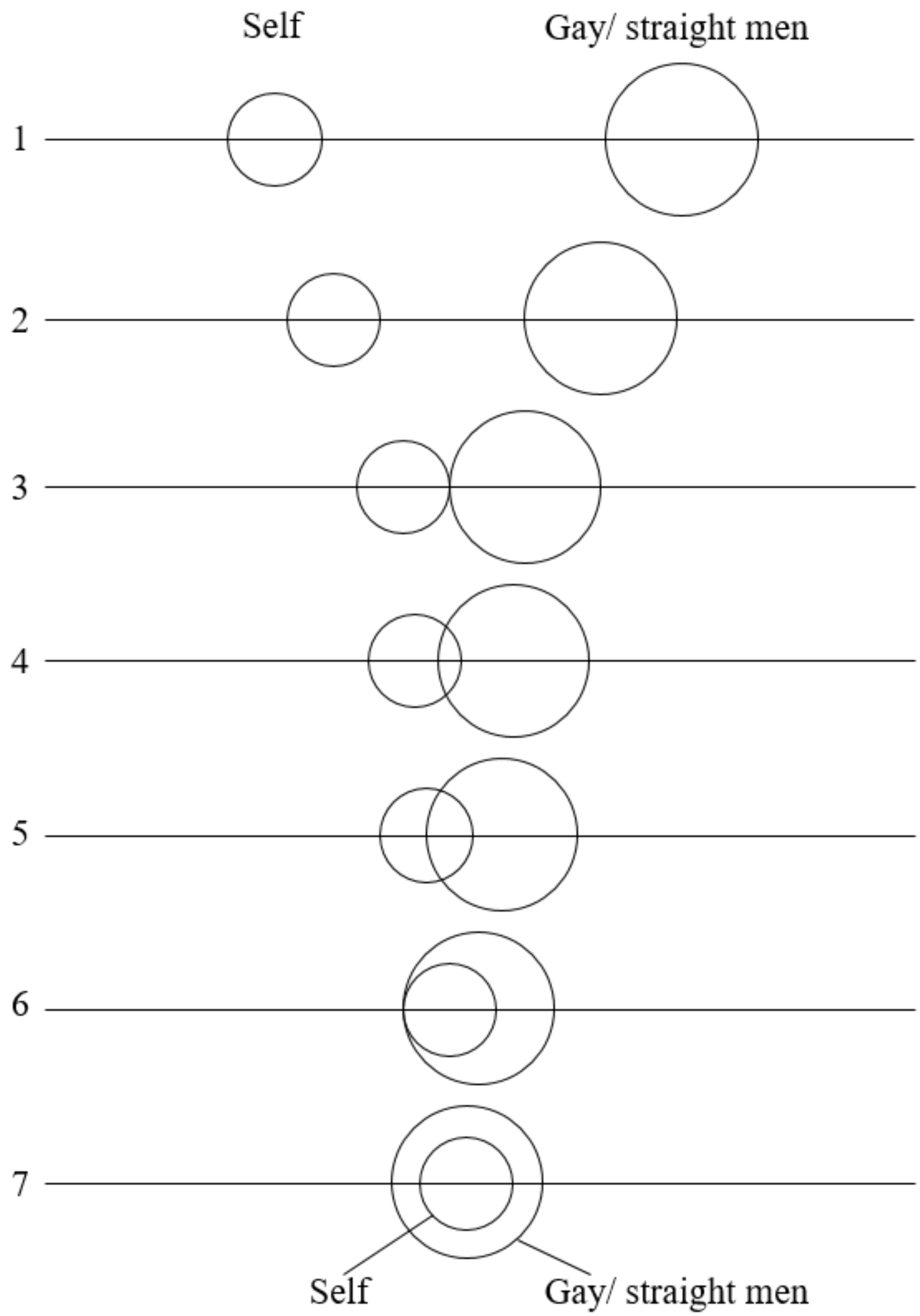
+We adapted the original CGNS scale for our study with regard to scale range (1-5), description of the scale points, and a few wordings. Moreover, we added the last item to our scale.

	I strongly disagree		Partly partly		I strongly agree
	1	*2*	*3*	*4*	*5*
I was a feminine boy.	○	○	○	○	○
As a child I was called a sissy by my parents and/or peers.	○	○	○	○	○
As a child I sometimes wished I had been born a girl rather than a boy.	○	○	○	○	○
As a child I preferred playing with girls rather than boys.	○	○	○	○	○
As a child I often felt that I had more in common with girls than boys.	○	○	○	○	○
As a child I usually avoided masculine clothing.	○	○	○	○	○
As a child I did not like competitive sports such as soccer or basketball.	○	○	○	○	○
* As a child I behaved like a typical boy.	○	○	○	○	○

*Items with asterisks had to be recoded (1 to 5, 2 to 4, 3 to 3, 4 to 2, 5 to 1).

Group affiliation (Affiliation to gay and straight men)

Item: "Please mark the picture that describes your relationship to the group of gay/straight men best."



Psycho-social identification

Collective Self-Esteem Scale (CSES by Luhtanen and Crocker, 1992).

	Strong rejection						Strong agreement
	1	*2*	*3*	*4*	*5*	*6*	*7*
Private-subscale							
*I often regret that I belong to this social group.	○	○	○	○	○	○	○
In general, I'm glad to be a member of the social group I belong to.	○	○	○	○	○	○	○
*Overall, I often feel that this social group is not worthwhile.	○	○	○	○	○	○	○
I feel good about the social group I belong to.	○	○	○	○	○	○	○
Identity-subscale							
*Overall, my group membership has very little to do with how I feel about myself.	○	○	○	○	○	○	○
The social group I belong to is an important reflection of who I am.	○	○	○	○	○	○	○
*The social group I belong to is unimportant to my sense of what kind of a person I am.	○	○	○	○	○	○	○
In general, belonging to this social group is an important part of my self-image.	○	○	○	○	○	○	○

*Items with asterisks had to be recoded (1 to 7, 2 to 6, 3 to 5, 4 to 4, 5 to 3, 6 to 2, 7 to 1).

Social environment

- Present social environment

Current contact to gay men (based on Hodson, Harry, & Mitchell, 2009).

Item: "Frequency of contact to gay men."

1	2	3	4	5	6	7
Never	Almost never	Seldom	Sometimes	Often	Very often	Always

Current contact to straight men.

Item: "Frequency of contact to straight men."

1	2	3	4	5	6	7
Never	Almost never	Seldom	Sometimes	Often	Very often	Always

Gender distribution of friends.

Item: "My circle of friends consists of..."

1	2	3	4	5	6	7
…women only	…women mainly	…somewhat more women	…women and men equally	…somewhat more men	…men mainly	…men only

Sexual orientation of male friends.

Item: “My female circle of friends consists of...”

1	2	3	4	5	6	7
… gay men only	… gay men mainly	…somewhat more gay men	… gay and straight men equally	…somewhat more straight men	… straight men mainly	… straight men only

- Childhood social environment

Contact to girls during childhood.

Item: “Contact to girls.”

1	2	3	4	5	6	7
Never	Very seldom	Seldom	From time to time	Often	Very often	Always

Contact to boys during childhood.

Item: “Contact to boys.”

1	2	3	4	5	6	7
Never	Very seldom	Seldom	From time to time	Often	Very often	Always

Voice self-assessment

	Strong rejection						Strongly agreement
	1	*2*	*3*	*4*	*5*	*6*	*7*
I believe other people think I speak in a masculine way.	○	○	○	○	○	○	○
I think I speak stereotypically straight.	○	○	○	○	○	○	○
*In other people's opinion, I speak stereotypically gay.	○	○	○	○	○	○	○
*Other people think I speak in a feminine way.	○	○	○	○	○	○	○
*In my opinion, I speak stereotypically gay.	○	○	○	○	○	○	○
I think I speak masculine.	○	○	○	○	○	○	○

*Items with asterisks had to be recoded (1 to 7, 2 to 6, 3 to 5, 4 to 4, 5 to 3, 6 to 2, 7 to 1).

Coming-out measures (Data collected for non-straight men only)

Coming-out extent.

Item: “Does your mother know about your sexual orientation?” Yes/ no.

Item: “Does your father know about your sexual orientation?” Yes/ no.

(The scores for parents’ items were adapted to the 7-point scale in order to get comparable values for an overall coming-out extent index: if neither mother nor father knew about their son’s sexual orientation, this was recoded as “1”; if only one parent knew about it, it was represented with “4”, and if both parents knew, a “7” was assigned.)

Item: “How many persons of the following groups, if existing, know about your sexual orientation?”

	None			Half of them			All of them
	1	*2*	*3*	*4*	*5*	*6*	*7*
Present straight friends	○	○	○	○	○	○	○
Acquaintances	○	○	○	○	○	○	○
Fellow students	○	○	○	○	○	○	○
Colleagues	○	○	○	○	○	○	○
Boss	○	○	○	○	○	○	○

Concealment Motivation(-subscale of LGBIS by Mohr and Kendra, 2011).

	Strong rejection					Strong agreement
	1	*2*	*3*	*4*	*5*	*6*
I keep careful control over who knows about my same-gender romantic relationships.	○	○	○	○	○	○
I prefer to keep my same-gender romantic relationships rather private.	○	○	○	○	○	○
My sexual orientation is a very personal and private matter.	○	○	○	○	○	○

Perceived discrimination experiences.

	Never	Once	Rarely	Sometimes	Often
	1	*2*	*3*	*4*	*5*
Other people act as if I was uncomfortable to them.	○	○	○	○	○
I was insulted or called names.	○	○	○	○	○
I feared something could happen to me.	○	○	○	○	○
Compared to other people I felt like I was treated unfair.	○	○	○	○	○
I was harassed.	○	○	○	○	○
I received poorer service in bars or restaurants.	○	○	○	○	○
In shops I was ignored or treated unfriendly.	○	○	○	○	○
I was threatened.	○	○	○	○	○
Other people treated me with less respect.	○	○	○	○	○
Other people ignored me.	○	○	○	○	○
I was treated less politely compared to other people.	○	○	○	○	○
I experienced disadvantages during renting a flat.	○	○	○	○	○
I lost my job or did not got one, because I’m lesbian or bisexual.	○	○	○	○	○

Explicit speech stereotypes

		Only gay						Only straight
		1	2	3	4	5	6	7
1	High	o	o	o	o	o	o	o
2	Slow	o	o	o	o	o	o	o
3	Variable	o	o	o	o	o	o	o
4	Loud	o	o	o	o	o	o	o
5	In dialect	o	o	o	o	o	o	o
6	Succinct	o	o	o	o	o	o	o
7	Unclear	o	o	o	o	o	o	o
8	Bright	o	o	o	o	o	o	o
9	Non-nasal	o	o	o	o	o	o	o
10	Blunt	o	o	o	o	o	o	o
11	Hoarse	o	o	o	o	o	o	o
12	Soft	o	o	o	o	o	o	o
13	Lots of pauses	o	o	o	o	o	o	o
14	Rather asking	o	o	o	o	o	o	o
15	Shrill	o	o	o	o	o	o	o
16	Strong	o	o	o	o	o	o	o
17	Lots of resonance	o	o	o	o	o	o	o
18	Clear	o	o	o	o	o	o	o
19	Hard	o	o	o	o	o	o	o
20	Breathy	o	o	o	o	o	o	o

		Only gay						Only straight
		1	2	3	4	5	6	7
21	Dark	o	o	o	o	o	o	o
22	Rather stating	o	o	o	o	o	o	o
23	Strained	o	o	o	o	o	o	o
24	Fast	o	o	o	o	o	o	o
25	Slack	o	o	o	o	o	o	o
26	Squeaky	o	o	o	o	o	o	o
27	Raspy	o	o	o	o	o	o	o
28	Sedate	o	o	o	o	o	o	o
29	Staccato	o	o	o	o	o	o	o
30	Monotonouos	o	o	o	o	o	o	o
31	Weak	o	o	o	o	o	o	o
32	Stagnant	o	o	o	o	o	o	o
33	Standard German	o	o	o	o	o	o	o
34	Nasal	o	o	o	o	o	o	o
35	Little pauses	o	o	o	o	o	o	o
36	Stretched	o	o	o	o	o	o	o
37	Low	o	o	o	o	o	o	o
38	Fluid	o	o	o	o	o	o	o
39	Quiet	o	o	o	o	o	o	o
40	Pointed	o	o	o	o	o	o	o

Manuscript 4

Kachel, S., Radtke, A., Skuk, V. G., Zäske, R., Simpson, A. P., & Steffens, M. C. (2017). Do They All Speak the Same? – Investigating Sexual Orientation Information Using Voice Averages*.* Invited to Revise and Resubmit in *PlosOne*.

Do they all speak the same? –

Investigating sexual orientation information using voice averages

Sven Kachel[1, 2*], André Radtke[1, 2], Verena G. Skuk[1, 3, 4], Romi Zäske[1, 3, 4], Adrian P. Simpson[1, 5], and Melanie C. Steffens[1, 2]

a) DFG Research Unit Person Perception, Institute of Psychology, Friedrich Schiller Universtity of Jena, Jena, Germany
b) Department of Social, Environmental, and Economic Psychology, University of Koblenz-Landau, Landau, Germany
c) Department for General Psychology and Cognitive Neuroscience, Institute of Psychology, Friedrich Schiller University of Jena, Jena, Germany
d) Department of Otorhinolaryngology, Jena University Hospital, Jena, Germany
e) Department of German Linguistics, Friedrich Schiller University Jena, Jena, Germany

*Corresponding author

E-mail: sven.kachel@uni-jena.de (SK)

Running title: Perceived sexual orientation after voice averaging

Abstract

While the perception of sexual orientation in voices often relies on stereotypes, it is unclear whether vocal stereotypes and accurate perceptions of sexual orientation are each signaled by acoustic cues common to speakers of a given group. To address these issues we used a novel voice morphing technique to create voice averages from voices that represent extremes of a given sexual orientation group either in terms of actual or perceived sexual orientation by others. Importantly, averaging preserves only those acoustic cues shared by the original speakers. 144 listeners judged the sexual orientation of twelve natural-sounding stimuli, each representing an average of five original utterances. Half of the averages were based on targets' self-ratings of sexual orientation: On a 7-point Kinsey-like scale, we selected targets who rated themselves strongly with a certain sexual orientation group. The other half were based on most salient ratings by others (i.e., on speech-related sexual-orientation stereotypes). Listeners judged sexual orientation from the voice averages with above-chance accuracy indicating that the perception of actual and stereotypical sexual orientation, respectively, are based on acoustic cues shared by speakers of the same group. Mean fundamental frequency and other typical acoustic parameters showed systematic variation depending on speaker gender and sexual orientation. Effects of sexual orientation were more pronounced for stereotypical voice averages than for those based on speakers' self-ratings, indicating that sexual-orientation stereotypes even exaggerate those differences present in most salient groups of speakers. Implications of our findings for stereotyping and discrimination are discussed.

Introduction

How do comedians and actors impersonate gay male voices? Based on sexual-orientation stereotypes, one might minimize nasality, whereas another might use a gay lisp, and a third might raise voice pitch. Similarly, while one listener might consider a high-pitched voice as a distinguishing characteristic between straight and lesbian women, another listener may choose pitch variability. When considering how sexual orientation is expressed in the voice, some people are convinced that there are multiple different ways of coding sexual orientation in voices. However, others think there is a common core of acoustic correlates of sexual orientation. The present research was designed to test the second assumption using an innovative voice morphing technique.

Previous research on acoustic and perceptual cues to sexual orientation in voices reflects the two implicit assumptions mentioned above: either there are multiple different ways of coding sexual orientation in voices or, alternatively, there is a common core of acoustic correlates of sexual orientation. In line with the first assumption that emphasizes intra-group diversity, speakers belonging to one sexual orientation group (e.g., straight men) use different acoustic cues to code their sexual orientation (see [1-4]). Studies following this approach assume, for instance, that one man indicates his straightness by using a low fundamental frequency while another applies a restricted vowel space in a specific situational context. According to this perspective, strictly speaking, on average there should be no acoustic differences between sexual orientation groups. In contrast, the second assumption considers speakers belonging to one sexual orientation group as rather homogeneous and systematically different from other sexual orientation groups with respect to specific acoustic parameters (e.g. [5]). Studies following this approach assume, for instance, that straight women use on average a higher fundamental frequency and a more expanded vowel space compared to lesbians and bisexual women. According to this perspective, there should be a uniform set of acoustic parameters for each sexual orientation group, leading to pronounced differences between sexual orientation groups. Listeners should be able to perceive these acoustic differences according to previous voice-based studies showing that straight people were rated as straighter compared to lesbians/gay men [6-9]. To the best of our knowledge, no voice-based perception study has investigated bisexual women and men as a separate category.

To answer the question whether systematic acoustic differences between sexual orientation groups do exist, previous studies have routinely aggregated individual acoustic parameters within groups and compared them between groups. However, findings are contradictory (see [10] for a recent review). That approach cannot determine if typical voice characteristics drive the perception of sexual orientation because different acoustic features could have influenced the perception of different individuals. To overcome this shortcoming, in the present research we created synthetic averages of natural utterances of the same expression by different speakers using voice morphing [11]. While voice morphing has not been used in research on sexual orientation, this powerful technique has been employed to study the perception of vocal gender and age [12, 13], emotional prosody [14], identity [15] or attractiveness [16]. Since voice morphing requires ample linguistic expertise and is technically challenging, most researchers have used simple vowel stimuli, rather than more complex sentence utterances, and have typically averaged voices of no more than two speakers at a time. Here we created high-quality voice samples across several speakers uttering a whole sentence. We thereby eliminated individual differences between speakers of the same sexual orientation group and created voices with characteristics typical of a given sexual orientation group.

Morphing results in one naturalistic sounding voice which is characterized by all acoustic features that are typical for and common to members of a specific sexual orientation group. Importantly, this voice can be presented to and judged by listeners, leading to a more direct assessment of the acoustic and perceptual correlates of sexual orientation.

If morphing cancels out varying intragroup cues to sexual orientation, listeners should not differently judge the sexual orientation of different sexual orientation groups. Alternatively, if sexual orientation information were still perceivable in voice averages, this would provide strong evidence that listeners who judge sexual orientation use a consistent set of acoustic cues typical for speakers of a given sexual orientation. This set of cues could represent valid signals to sexual orientation and/or speech stereotypes held by listeners. According to some of the oldest theories of stereotypes, stereotypes contain a "kernel of truth" (see [17]), but exaggerate differences between social groups [18]. Regarding speech stereotypes there is evidence for a gender difference: speech stereotypes regarding gay and straight men are more pronounced than regarding lesbians and straight women [19]. We tested differences between social groups by using single voices of speakers who identified their sexual orientation themselves (i.e., self-rated sexual orientation). In addition, we tested stereotypes by selecting recordings of speakers who were perceived as very typical for one sexual orientation group in previous studies (i.e., sexual orientation rated by others). There were two other reasons for using these two different sets of speakers. First, voices that were previously rated by others as very typical were assumed to show pronounced effects on perceived sexual orientation after averaging. This is important because we expected rather small effects. Second, self- and other-rated sexual orientation may rely on different sets of acoustic cues. We analyzed acoustical parameters commonly investigated in sociophonetic studies on sexual orientation in order to determine which of them differ between sexual orientation groups. These include mean f0 [20-23], vowel space expansion, vowel space shift, and F1 and F2 values in single vowels [5,7,24,25]. Hence, we will be focusing on intonational features and vowel space characteristics in the present study.

The present study

The aim of the present study was to test whether the same typical voice characteristics drive the perception of sexual orientation of different speakers. In order to increase the probability of finding common acoustic features, if these exist at all, we created maximally different voice averages using an innovative voice morphing technique. We created twelve natural-sounding stimuli, each representing an average of five original utterances. Half of the averages were based on targets' self-ratings of sexual orientation (as 1 = lesbian/gay, 4 = bisexual women/men, or 7 = straight women/men). The other half were based on most salient ratings by others (i.e., on speech-related sexual-orientation stereotypes). Listeners judged sexual orientation from the voice averages.

Based on the idea that different speakers use a consistent set of acoustic cues to signal sexual orientation, and that listeners can use that set of cues, we tested the following hypotheses:

1. Gay/lesbian speakers are judged as less straight than bisexual speakers, who are in turn judged as less straight than straight speakers.
2. These differences are found both for female and male speakers but more pronounced for male speakers because of stronger stereotypes regarding male than female sexual orientation.

3. In line with the "kernel of truth" hypothesis of stereotyping, effects should be more pronounced for sexual orientation groups based on ratings by others (i.e, for stereotypes) than for sexual orientation groups based on self-ratings.
4. In line with speech stereotypes, straight women's voice average should show higher mean f0 than bisexual women's voice average that in turn should be higher than lesbians' voice average when based on ratings by others (vice versa for male voice averages).

Method

Ethics statement

The experiment was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the University of Jena, Faculty of Social and Behavioral Sciences (approval number FSV 12/02).

Speakers

Initially, 111 (57 female, 54 male, M_{age} = 24.08, SD_{age} = 2.55, age range 19 – 30 years) speakers were recorded under standardized conditions in a sound-treated room [10,26]. From a total of twenty different sentences, the target sentence for this study was selected: *Der Tag ist sehr lang geworden.* (/deːr taːk ɪst zeːr laŋ ɡəvɔrdən/; "It has been a very long day."). Each sentence was spoken three times in a neutral way. This sentence was chosen because it contains examples of features that have typically found to differ based on gender and sexual orientation, e.g., vowel space dimensions in the range of front-back, close-open vowels; sibilant measures for /s z/; VOT in initial fortis plosive of *Tag*. The sentence was also considered to be suitable as a basis for creating averages of several utterances: it is segmentally relatively simple and we would not expect much inter-individual variation in the segmental realization.

All speakers provided information on their sexual orientation on a 7-point *Kinsey-like scale* (1, exclusively lesbian/gay; 4, bisexual, and 7, exclusively straight) and on a 7-point scale separating sexual orientation towards men and women. The latter included physical attraction, sexual fantasies, romantic emotions, and sexual interaction (for details see [27]). The rating difference between items belonging to sexual orientation towards men and women served as a S*exual Orientation Index,* with positive scores indicating stronger sexual orientation toward men and negative scores a stronger sexual orientation toward women.

Selection of speakers for the present study

To create voice averages, single voice recordings were selected according to four criteria. First, recordings of five speakers were selected to create one voice average (e.g., five voices of women who were rated by others as bisexual). Second, we selected speakers with the most typical sexual orientation scores regarding self-ratings or ratings by others in order to maximize acoustic information indicating sexual orientation (e.g., extreme ratings for lesbian/gay and straight speakers and intermediate ratings for bisexual speakers). Third, due to different numbers of speakers possibly available for each voice average (e.g., few men rated themselves as bisexual but a lot were rated as bisexual by others), we preferred selection for voice averages where small numbers of speakers (vs. high numbers) were expected. Fourth, in order to avoid redundant signals and to maximize contrasts between voices, a given voice was not selected for more than one voice average.

Two sets of speakers were selected for the purpose of the current study. A first set of 30 speakers was selected based on the speakers' self-ratings (15 female, 15 male; 5 in each of the groups: lesbian/gay, bisexual, or straight, respectively). The primary differentiation was done according to the Kinsey-like scale because it provides a clearly identity-based assignment as either exclusively lesbian/gay or straight, or very close to the scale's center (i.e., bisexual). As a secondary differentiation, we used the more fine-grained Sexual Orientation Index. Whenever more than five speakers had the same Kinsey-like score, those were chosen whose Sexual Orientation Index was most typical for a given sexual orientation group. In this way, most salient groups of speakers were created that differed maximally regarding self-rated sexual orientation.

A second set of 30 speakers was selected based on sexual orientation rated by others (15 female, 15 male; five in each of the groups: lesbian/gay, bisexual, or straight, respectively). From the whole pool of 111 speakers, 18 lesbians, gay men (Kinsey-like scores 1-2), straight women and men (Kinsey-like scores: 6-7) each were randomly chosen to be rated as either lesbian/gay or straight based on sentence recordings by two independent samples of a total of 216 participants (121 women, 95 men, $M_{age} = 28.12$, $SD_{age} = 10.53$, age range 18 – 71 years), in two different studies ([28] and an unpublished Master thesis [29]). Based on these ratings, we computed mean ranks and mean rank differences between the two independent rater samples. Selection for the "ratings by others" condition was primarily based on mean ranks. Whenever two speakers showed the same mean ranks, those with smaller mean rank differences were preferred because raters' perceptions regarding that speaker were more similar. In this way, maximally different groups regarding speech stereotypes were created. Self-ratings and ratings by others for each voice average are listed in Table 1.

Table1. Overview of sexual orientation ratings and fundamental frequency features, vowel space characteristics, and voice quality measures for all voice averages (in Hz).

	5-voice female averages						5-voice male averages					
	Based on self-ratings			Based on ratings by others			Based on self-ratings			Based on ratings by others		
	L	B	S	L	B	S	G	B	S	G	B	S
Age	23.20	23.00	23.00	23.20	25.00	23.20	25.80	24.60	25.60	24.80	24.20	23.40
Self-ratings of sexual orientation												
Kinsey-like scale[a]	1.00	4.00	7.00	2.80	5.20	4.80	1.00	4.40	7.00	2.60	2.60	4.80
Sexual orientation index[a]	-5.25	.00	4.40	-2.30	2.25	.85	5.65	-.10	-5.75	3.45	2.45	-1.75
Ratings of sexual orientation by others												
Perceived straightness[b]	.80		.88	.62	.80	.90	.65	.70	.66	.31	.66	.87
Mean ranks[b]	17.50		26.83	3.40	15.20	31.50	14.00	16.5	19.83	3.00	13.75	32.00
Mean rank differences[b]	3.67		-.33	-.40	-1.20	-.60	6.00	13.00	.33	.00	.60	-1.60
Fundamental frequency features												
f0 mean	209	190	198	176	171	206	122	113	106	117	109	106

f0 *SD*	10	10	19	20	17	21	15	11	10	12	10	17
f0 2.5th percentile	191	175	168	134	146	174	99	95	85	100	93	81
f0 97.5th percentile	235	210	231	203	203	239	144	134	126	137	125	135
Vowel space characteristics												
F1 mean	624	612	582	562	600	625	520	559	545	526	545	500
F2 mean	1403	1436	1422	1363	1408	1408	1306	1238	1173	1292	1269	1183
Vowel space dispersion	359	410	387	355	405	360	371	394	368	345	398	328

Abbreviations: L = lesbians, G = gay men, B = bisexual wo/men, S = straight wo/men, SO = sexual orientation, SD = standard deviation, **Kinsey-like scale** ranged from 1 – "exclusively lesbian/gay" via 4 – "equally lesbian/gay and straight" to 7 – "exclusively straight". **Sexual orientation index** ranged from -7 – "sexually oriented towards women only" via 0 – "sexually oriented towards women and men" to 7 – "sexually oriented towards men only". **Perceived straightness** indicates mean relative numbers of perceptions as straight in both pre-ratings ranging from 0 – "judged as straight by 0%" to 1 – "judged as straight by 100%". **Mean ranks** ranged from 3 – "voices were on average located at the lesbian/gay end of the perceived sexual orientation distribution" via 18.50 – "voices were on average located at the bisexual area of the perceived sexual orientation distribution" to 34 – "voices were on average located at the straight end of the perceived sexual orientation distribution". **Mean rank differences** close to 0 indicate that voices were rated as very similar on average in both pre-studies. Note that we did not depict standard deviations for self-ratings and ratings by others because of small sample sizes (*n* = 5 for each voice average).

[a]*n* = 5. [b]*n* for calculating mean ranks and mean rank differences ranged from 0 for women who rated themselves as bisexual to 4 for men who rated themselves as gay, because not every speaker who was selected for voice averaging based on self-ratings was selected for ratings by others

Stimuli

Preparation. For each of the selected 60 speakers, we chose the utterance (out of three) judged to have a segmental make-up most appropriate for morphing. Whenever signal quality was comparable across the three repetitions of the sentence, we selected the one that was contained in the pre-ratings (*n* = 37). Whereas this could be considered a methodological shortcoming, we argue that straightness should be invariant for the same speaker of the same utterance in the same situation. Moreover, voice averaging would eliminate such subtle intra-group differences. 50 ms of silence was inserted before and after each utterance. Artefacts, such as clicks, were manually deleted before morphing. Additionally, selected recordings were root-mean-square normalized at 70 dB SPL using Praat [30].

Voice averaging. Five-voice averages were created for every condition (sexual orientation group: lesbian/gay vs. bisexual vs. straight x speaker gender: female vs. male x rating basis: self- vs. other-ratings). Voice averaging was performed using the novel n-way morphing approach of the speech analysis, modification and resynthesis framework TANDEM-STRAIGHT [11], which creates a single average voice based on the independent representations of source- and filter information of *n* voices. Voice averaging requires a manual annotation of time- and frequency-anchors of corresponding key-features in the spectrograms of the original utterances (e.g., onsets and offsets of phonemes, vowel transitions). For the present sentence-morphs a total of 37 time-anchors was required per voice sample; frequency anchors were placed at center frequencies of formants (F1-F4, where detectable) at each time-anchor location.

Listeners

Overall, 237 participants took part in the online listening experiment and 155 completed it. Data from 10 listeners were excluded because they were non-native German speakers, reported hearing impairments, or had participated in previous studies on the judgement of sexual orientation based on voices. Moreover, we excluded data from one person whose gender was described as neither male nor female because we wanted to test for possible effects of listener gender.

Accordingly, data of 144 listeners (110 female, 34 male, $M_{age} = 23.13$, $SD_{age} = 3.67$, age range 18 – 43 years) were analyzed. The experiment was announced via a students-only mailing list at the University of Koblenz-Landau as investigating the prejudice that people's sexual orientation could be recognized based on their voices.

Design and procedure

A total of twelve voice averages were available as stimuli: three sexual orientation groups (lesbian/gay vs. bisexual vs. straight), two speaker genders (female vs. male), and two types of ratings (self-rated vs. rated by others). After reporting sociodemographic and other relevant characteristics, listeners heard a single test voice (that was not part of any voice average) and were instructed to select a comfortable sound level. During the experiment, the listeners were asked to judge each voice average on a 7-point scale ranging from 1 "very lesbian"/"very gay" via 4 "bisexual" to 7 "very straight" by mouse-click. Listeners were told that the stimuli displayed the whole range of very lesbian/gay to very straight sounding voices. Male and female voice averages were presented in separate blocks, to avoid switching between gender-related standards of voice perception. Listeners were randomly assigned to one of two block orders (i.e., 76 listeners heard female voice averages first, the other 68 heard male voice averages first). Within each block, stimuli were randomly presented for every listener. Mean perceived sexual orientation was computed for each voice average with higher scores indicating higher perceived straightness.

Acoustic analysis

For each of the voice averages typical acoustical features were measured using Praat [30]. These included f0 measures (mean, *SD*, 2.5th and 97.5th percentile indicating lower and upper f0 boundary) and mean formant frequencies (F1-F2) of three vowels (corresponding to the inner section, i.e. 25th to 75th percentile: /a:/ in "Tag", /ɪ/ in "ist", and /ɔ/ in "geworden"). F0 was tracked every 10ms for the whole duration of each voice average. In total, for the voiced parts of all voice averages, 1306 data points were available which is more than appropriate for a statistical mean f0 comparison between voice averages. F1 and F2 were tracked every 6ms during half of a vowel's duration centered around the vowel midpoint. From the F1 and F2 values, we computed vowel space dispersion and vowel space shift. Vowel space dispersion is defined as the mean Euclidian distance of the three vowels from the center of the vowel triangle [31]; vowel space shift is indicated by mean F1 and F2 across the three vowels. Because for each voice average only one data point for vowel space dispersion, mean F1 and F2, and approximately 5 data points for F1 and F2 of every single vowels were available, statistical comparison between voice averages is inappropriate. However, we provide descriptive findings for vowel space characteristics.

Results

Perception of sexual orientation

A 3 x 2 x 2 x 2 x 2 ANOVA was performed on perceived straightness, with three within-subject factors (speaker sexual orientation, speaker gender, and rating basis), and two between-subject factors (listener gender and block order). This analysis did not reveal any main effects or interactions involving listener gender or block order (all $p > .056$).

The analysis revealed main effects of speaker sexual orientation, $F(2, 280) = 125.53$, $p < .001$, $\eta_p^2 = .47$, and rating basis, $F(1, 280) = 45.65$, $p < .001$, $\eta_p^2 = .25$. Main effects were qualified by three first-order interactions of speaker sexual orientation x rating basis, $F(2, 280) = 23.69$, $p < .001$, $\eta_p^2 = .16$, speaker sexual orientation x speaker gender, $F(2, 280) = 9.16$, $p < .001$, $\eta_p^2 = .06$, and speaker gender x rating basis, $F(1, 280) = 8.64$, $p = .004$, $\eta_p^2 = .06$, all other effects: $F(2, 280) < 2.01$, $p \geq .136$, $\eta_p^2 \leq .01$. For a detailed analysis of which voice averages differ from each other, we applied simple-effects tests with Bonferroni adjustment.

Fig 1a shows the interaction of speaker sexual orientation x rating basis. Straight voice averages were perceived as straighter compared to bisexual voice averages that were, in turn, perceived as straighter than lesbian/gay voice averages; that was true for ratings by others, $F(2, 139) = 116.53$, $p < .001$, $\eta_p^2 = .63$ (all pair-wise ps $\leq .001$), and self-ratings, $F(2, 139) = 32.33$, $p < .001$, $\eta_p^2 = .32$ (all pair-wise ps $\leq .021$), supporting Hypothesis 1. According to effect sizes the differences between sexual orientation groups were more pronounced for voice averages based on ratings by others (i.e., sexual-orientation stereotypes) than self-ratings (i.e., actual differences between most salient groups), in line with Hypothesis 3. Moreover, only lesbian/gay voice averages differed regarding the rating basis: Self-rated lesbian/gay voice averages were perceived as straighter than those rated by others, $F(1, 140) = 89.55$, $p < .001$, $\eta_p^2 = .39$. In bisexual and straight voice averages rating basis had no effect, both $F(1, 140) \leq 2.41$, both $p \geq .123$, both $\eta_p^2 \leq .02$. This finding partially supports our initial assumption that even after voice averaging, effects of perceived sexual orientation were more pronounced when single voices were previously rated as very typical for one sexual orientation group by others compared to self-ratings. In other words, speech stereotypes of sexual orientation exaggerate existing group differences.

Fig 1b shows the same pattern as Fig 1a but illustrates the interaction of speaker sexual orientation and speaker gender. Straight voice averages were perceived as straighter than bisexual voice averages, which, in turn were perceived as straighter in contrast to lesbian/gay voice averages; that was true for male, $F(2, 139) = 85.48$, $p < .001$, $\eta_p^2 = .55$ (all pair-wise *p*s < .001) and female speakers, $F(2, 139) = 44.61$, $p < .001$, $\eta_p^2 = .39$ (all pair-wise *p*s < .001), in line with Hypothesis 2. According to effect sizes, differences in perceived sexual orientation were more pronounced for male than female sexual orientation groups. Hence, independent of rating basis and speaker gender, straight voice averages were perceived as straighter than bisexual voice averages, which, in turn, were perceived as straighter than lesbian/gay voice averages. Furthermore, gay voice averages were perceived as less straight compared to lesbian voice averages, $F(1, 140) = 8.70$, $p = .004$, $\eta_p^2 = .06$. No gender difference occurred for bisexual voice averages, $F(1, 140) = 1.74$, $p = .190$, $\eta_p^2 = .01$, and straight men's voice averages were perceived as straighter compared to straight women's voice averages, $F(1, 140) = 4.13$, $p = .044$, $\eta_p^2 = .03$. Hence, fully supporting Hypothesis 2, differences were more pronounced for male than female speakers.

Fig 1c illustrates the interaction of speaker gender and rating basis. Voice averages based on self-ratings were perceived as straighter compared to voice averages based on ratings by others; this was true for female, $F(1, 140) = 40.09$, $p < .001$, $\eta_p^2 = .22$, and male voice averages, $F(1, 140) = 11.53$, $p = .001$, $\eta_p^2 = .08$, but differences were slightly more pronounced for female than male voice averages according to effect sizes. Within female voices, $F(1, 140) = 2.75$, $p = .100$, $\eta_p^2 = .02$, and male speakers, $F(1, 140) = 3.54$, $p = .062$, $\eta_p^2 = .03$, voice averages based on self-ratings vs. ratings by others did not differ significantly.

Summing up, sexual orientation information was still perceived after voice averaging of maximally different groups: Straight voice averages were perceived as straighter compared to bisexual voice averages that were still perceived as straighter compared to lesbian/gay voice averages. The order was unaffected by rating basis and speaker gender. However, the differences in perceived sexual orientation between sexual orientation groups were more pronounced in ratings by others (vs. self-ratings) and in male speakers (vs. female speakers). Thus, most salient members of one sexual orientation group seem to share a consistent set of acoustic cues that listeners use to judge their sexual orientation, but speech stereotypes exaggerate existing group differences.

Acoustic correlates of sexual orientation

Having demonstrated that a consistent set of acoustic cues drives the perception of extreme groups' sexual orientation, we asked which acoustic characteristics differ between sexual orientation groups and may thus lead to differences in perceived straightness. Acoustic values for fundamental frequency features (mean f0, f0 *SD*, 2.5th percentile, 97.5th percentile) and vowel space characteristics (mean F1, mean F2, vowel space expansion) for every voice average can be seen in Table 1. On a descriptive level, 22 out of 32 acoustic measures differed according to speech stereotypes when comparisons were based on voice averages of straight and lesbian/gay speakers (excluding bisexual voice averages). The number of stereotype-congruent differences were twice as high for women's voice averages based on ratings by others ($n = 7$) compared to voice averages based on self-ratings ($n = 3$), whereas rating basis did not affect number of differences for men's voice averages (n = 6 for self-ratings and ratings by others). Note that statistical tests were performed only for mean f0, as this was the only parameter with a sufficient number of data points for robust analysis (i.e. one data point per 10 ms of speech).

A 3 x 2 x 2 between-subjects ANOVA yielded main effects of speaker sexual orientation, $F(2, 1294) = 58.26$, $p < .001$, $\eta_p^2 = .08$, speaker gender, $F(1, 1294) = 9155.80$, $p < .001$, $\eta_p^2 = .88$, and rating basis, $F(1, 1294) = 122.79$, $p < .001$, $\eta_p^2 = .08$. Main effects were qualified by interactions of speaker sexual orientation x rating basis, $F(2, 1294) = 64.91$, $p < .001$, $\eta_p^2 = .09$, speaker gender x speaker sexual orientation, $F(2, 1294) = 97.66$, $p < .001$, $\eta_p^2 = .13$, and speaker gender x rating basis, $F(1, 1294) = 56.89$, $p < .001$, $\eta_p^2 = .04$, which, in turn, were qualified by a higher-order interaction of speaker sexual orientation x speaker gender x rating basis, $F(2, 1294) = 41.13$, $p < .001$, $\eta_p^2 = .06$ (see Fig 2). For a detailed analysis regarding which voice averages differed from each other, we applied simple-effects tests with Bonferroni adjustment.

Mean f0 values for all pairwise comparisons between voice averages differed significantly, $F(2, 1294) \geq 4.19$, $p \leq .041$, $\eta_p^2 \geq .01$ (all pairwise ps $\leq .041$), except for the comparisons between voice averages of bisexual males rated by others, self-rated straight males, and straight males rated by others, $F(2, 1294) < 1$ (all pair-wise ps $\geq .49$). Hence, for self-rating averages, straight men were lower in mean f0 than bisexual men who, in turn, had lower mean f0 values compared to gay men. In contrast, for averages based on ratings by others, both straight and bisexual men were comparable in mean f0 but were lower than gay men. Regarding female voice averages, the mean f0 pattern was as follows. Voice averages of bisexual women showed lowest mean f0 compared to lesbian and straight voice averages, both in the self-rating and rating by others condition. In the case of self-ratings, voice averages of bisexual women showed values more similar to voice averages of straight women with voice averages of lesbians having the highest mean f0. However, in the case of ratings by others, voice averages of bisexual women showed values more similar to voice averages of lesbians, with voice averages of straight women having the highest mean f0. Hence, mean f0 patterns according to ratings by others were in line with speech stereotypes (straight women displaying higher voice pitches than lesbians) whereas the mean f0 pattern according to self-ratings contradicts these stereotypes. This conclusion was also supported when looking at the differences in mean f0 according to self-ratings vs. ratings by others. The largest difference in rating basis appeared in lesbians' voice averages ($d = 2.09$) followed by bisexual women's voice averages ($d = 1.43$; all other: $d < .05$): Mean f0 for the voice average for self-rated lesbians was ca. 35 Hz higher than for those who were rated as lesbian by others. By implication, it is unclear on which basis the sexual orientation of self-rated lesbians was accurately determined.

Discussion

Here we used acoustic morphing to eliminate inter-individual acoustic differences between voices of speakers, either who identified themselves strongly with a certain sexual orientation, or who listeners strongly perceived of as having a certain sexual orientation. Overall, voice averages of straight speakers were perceived as straighter than bisexual speakers who, in turn, were perceived as straighter compared to lesbian/gay speakers, irrespective of speaker gender and whether speaker sexual orientation was self-rated or had been rated by others (i.e., was based on speech stereotypes). Thus, sexual orientation information can still be perceived in voice averages derived from speakers who represented the most extreme or the most salient examples of a given sexual orientation. Whereas differences in perceived sexual orientation were more pronounced for male than female sexual orientation groups, the effect was clearly present for female voices, too. This finding allows two interpretations: First, listeners who judged sexual orientation used a consistent set of acoustic cues typical for most salient speakers of a given sexual orientation, and second, there are systematic acoustic differences between extreme sexual orientation groups.

Because we assumed perceived sexual orientation differences between voice averages to be rather small, if present at all, we tried to maximize sexual orientation information and to create maximally different voice averages by selecting those speakers for voice averaging who showed extreme scores on self-rated and other-rated sexual orientation. Thus, there was a systematic selection of voices chosen for averaging regarding a certain sexual orientation group. As another possible limitation of the present study, speakers were not blind to the purpose of the study. We can therefore not rule out some degree of self-selection among the speakers. Specifically, an increased number of lesbian/gay speakers who felt comfortable with their sexual orientation may have contributed to more stereotypical speaking patterns overall. Note however, that the issue of representativeness would not be resolved even if we had included all speakers from the initial sample to create the voice averages. Creating representative samples of speakers with respect to sexual orientation is an unrealistic undertaking [32-34]. Importantly, our findings do not imply that the sexual orientation of less extreme groups, nor of individuals, can be judged accurately. Instead, they demonstrate that maximally different groups of speakers, and especially groups stereotyped as most salient based on voice samples, use some common acoustic features. Our study should be taken as a starting point for future research that could test whether less salient sexual orientation information is still preserved after voice averaging, that is, when randomly selecting lesbians and gay men.

Group stereotypes have been regarded as exaggerations of existing differences between social groups (e.g., [18]). The most important interaction that we found was between speaker sexual orientation and rating basis. This interaction suggests that people hold stereotypes of lesbian and gay speech more than of bisexual or straight speech: Whereas self-ratings and other-ratings converged on very similar findings regarding bisexual or straight targets, lesbian/gay voice averages based on other-ratings were perceived as much more lesbian/gay than those based on self-ratings. In other words, the voice average formed by the five speakers who most clearly self-identified as lesbian/gay was judged as rather straight, whereas the voice average that consisted of the five speakers who sounded most lesbian/gay was indeed judged as most lesbian/gay. Hence, this finding confirms our assumption that perceived sexual orientation differences were more pronounced in voice averages of speakers whose sexual orientation had been previously rated by others, confirming for speech stereotypes regarding sexual orientation according to Allport's [18] old assumption that stereotypes are exaggerations.

Findings on mean f0 as a crucial and probably most influential acoustic marker of gender and sexual orientation of voice averages showed that ratings by others are in line with the idea of stereotype-driven perceptions at least for female speakers: Women who were rated as lesbian or bisexual showed lower mean f0 values than those who were rated as straight. In contrast, self-rated lesbians produced highest mean f0 values compared to self-rated bisexual and straight women. This mean f0-pattern challenges a core approach in research on acoustic cues of sexual orientation (e.g., [5]): Lesbians and bisexual women (and later gay and bisexual men as well) were lumped together because they showed no differences (supposedly in acoustic parameters). By contrast, our results support the idea that bisexual women were acoustically different from lesbians and straight women. The same is true for self-ratings of men: Bisexual men were in-between gay and straight men. However, according to ratings by others, bisexual and straight men could be lumped together. Thus, we recommend treating bisexual people as a distinct sexual orientation group, as we did for voice-based perception for the first time (for face-based perception of bisexual people, see [35]). Whereas differences in perceived sexual orientation were more pronounced for male than female speakers, differences in mean f0 were more pronounced for female speakers than male speakers. Hence, acoustic parameters other than mean f0 likely affect differences in perceived sexual orientation groups for male speakers. Future research should use single-parameter morphs to investigate this question.

Our findings second previous research indicating the use of stereotypical information when judging others' sexual orientation (see [36]). When assuring high stereotypical information of voice averages (i.e., selecting single voices for averages based on sexual orientation ratings by others), differences between sexual orientation groups were more pronounced than when assuring lower stereotypical information (i.e., selection based on speakers' self-ratings). Moreover, differing mean f0 patterns provided further evidence: Mean f0 of voice averages with sexual orientation rated by others showed an overall stereotypical pattern whereas voice averages with self-rated sexual orientation did not.

Previous studies applying voice morphing techniques used simple stimuli, such as vowel-consonant-vowel syllables. In the present study, more complex sentence stimuli were morphed for the first time. Thus, our findings reflect acoustic and perceptual correlates of sexual orientation based on stimuli with increased ecological validity. Hence, we recommend the use of sentence stimuli in future voice morphing studies.

It has to be mentioned that the present study was carried out in a particular culture, time, and language, and voluntary participants provided the speech samples that we used as the basis of our voice averages. Manners of speaking may vary with the language used, with role models, and with the sample investigated. Thus, the generalizability of our study's findings needs to be tested by future research that may rely on the technological advances that we used for the first time.

Conclusion

The present study set out to answer the question whether there is a common core of vocal expressions of sexual orientation in maximally different groups, or whether different individuals use different means to express their sexual orientation. To the best of our knowledge, this study was the first to use voice averages of complex stimuli (i.e., sentence-length utterances) to test this. We found an effect of sexual orientation across conditions, in line with the idea that the selected groups of most salient speakers used similar ways of expressing sexual orientation. Our findings do not allow conclusions regarding single speakers or less extreme groups (e.g., with sexual orientation ratings of 2 or 6 on a 7-point scale). Additionally, the observed effect of sexual orientation was larger when voice averages were based on perceivers' ratings than on self-ratings, demonstrating that stereotypes of gay or lesbian manners of speaking are exaggerations of the differences that truly exist between speakers.

Acknowledgments

Substantial contributions to the conception or design of the work: SK, AR, VGS, RZ, APS; substantial contribution to the acquisition and analysis of the data: SK, AR, VGS, MCS; interpretation of data for the work: SK, VS, RZ, APS, MCS; Drafting the paper and revising it: SK, AR, VS, RZ, APS, MCS; Final approval of the version to be published: SK, AR, VS, RZ, APS, MCS.

The current research was funded by a grant from the German Research Foundation to the last two authors (DFG STE 938/10-2; FOR 1097). We thank Jörg Dreyer, Dr. Susanne Fuchs and Dr. Stefanie Jannedy for providing the phonetic laboratory at the Zentrum für Allgemeine Sprachwissenschaft in Berlin, we thank Rebecca Aust, Susanne Friedrich, Alexander Makosch, Sonja Gaukel, Felicia Schuld, Felix Sokoll, Karoline Nestler, and Anna Therasa Prokein for help with data collection, data pre-processing, and formatting of the manuscript. The authors declare no conflicts of financial, personal, or other interests.

References

1. Zwicky, A. M. (1997). Two lavender issues for linguists. In A. Livia & K. Hall (Eds.), *Queerly phrased. Language, gender, and sexuality*, (pp. 21-34). New York: Oxford University Press.

2. Zimman, L. (2010). Female-to-male transsexuals and gay-sounding voices: A pilot study. Colorado Research in Linguistics (University of Colorado, Boulder), Volume 22.

3. Eckert, P., & Podesva, R. J. (2011). Sociophonetics and sexuality: Towards a symbiosis of sociolinguistics and laboratory phonology. *American Speech, 86,* 6–13.

4. Munson, B. (2011). Lavender Lessons Learned: Or, What Sexuality Can Teach Us About Phonetic Variation. *American Speech, 86,* 14–31.

5. Pierrehumbert, J. B., Bent, T., Munson, B., Bradlow, A. R., & Bailey, M. J. (2004). The influence of sexual orientation on vowel production (L). *J Acoust Soc Am, 116,* 1905–1908.

6. Bailey, J. M. (2003). *Gender Nonconformity and Sexual Orientation,* paper presented at the annual meeting of the Human Behavior and Evolution Society, 4–8 June, 2003.

7. Munson, B., McDonald, E. C., DeBoe, N. L., White, A. R. (2006). The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *J Phon, 34,* 202–240. DOI: 10.1016/j.wocn.2005.05.003.

8. Rieger, G., Linsenmeier, J. A. W., Gygax, L., Garcia, S., & Bailey, J. M. (2010). Dissecting 'gaydar': Accuracy and the role of masculinity–femininity. *Arch Sex Behav, 39,* 124–140. DOI: 10.1007/s10508-008-9405-2.

9. Valentova, J. V., & Havlíček, J. (2013). Perceived sexual orientation based on vocal and facial stimuli is linked to self-rated sexual orientation in czech men. *PLoS ONE, 8,* 1–6. DOI: 10.1371/journal.pone.0082417.

10. Kachel, S., Simpson, A. P., & Steffens, M.C. (2017a). Do I sound gay – Acoustic correlates of actual and perceived sexual orientation and masculinity/femininity in men's speech. Unpublished manuscript.

11. Kawahara, H., Morise, M., Banno, H., & Skuk, V.G. (2013). *Temporally variable multi-aspect N-way morphing based on interference-free speech representations.* Paper presented at the Signal and Information Processing Association Annual Summit and Conference (APSIPA), Asia Pacific.

12. Zäske, R., Skuk, V. G., Kaufmann, J. M., & Schweinberger, S. R. (2013). Perceiving vocal age and gender: An adaptation approach. *Acta Psychol (Amst), 144*, 583-593.

13. Skuk, V. G., & Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *J Speech Lang Hear Res, 57*, 285-296.

14. Bestelmeyer, P. E. G., Rouger, J., DeBruine, L. M., & Belin, P. (2010). Auditory adaptation in vocal affect perception. *Cognition*, 117(2), 217-223.

15. Zäske, R., Schweinberger, S. R., & Kawahara, H. (2010). Voice aftereffects of adaptation to speaker identity. *Hear Res, 268*(1-2), 38-45.

16. Bruckert, L., Bestelmeyer, P., Latinus, M., Rouger, J., Charest, I., Rousselet, G. A., . . . Belin, P. (2010). Vocal Attractiveness Increases by Averaging. *Curr Biol, 20*(2), 116-120.

17. Prothro, E. T., & Melikian, L. H. (1955). Studies in stereotypes: V. Familiarity and the kernel of truth hypothesis. *J Soc Psychol, 41*, 3-10. DOI: 10.1080/00224545.1955.9714248

18. Allport, G. W. (1954). *The nature of prejudice*. Massachusetts: Addison-Wesley.

19. Piccolo, F. (2008). Listener-identified phonetic correlates of gay-, lesbian- and straight-sounding speech. Unpublished PhD thesis.

20. Linville, S. E. (1998). Acoustic Correlates of Perceived versus Actual Sexual Orientation in Men's Speech. *Folia Phoniatr Logop,* 50, 35–48.

21. Baeck, H., Corthals, P., & van Borsel, J. (2011). Pitch characteristics of homosexual males. *J Voice*, *25*, 211–214. DOI: 10.1016/j.jvoice.2010.10.019.

22. Camp, M. (2011). *Japanese lesbian speech: Sexuality, gender identity, and language* (PhD thesis). Retrieved from ProQuest Dissertations and Theses, 1–216.

23. van Borsel, J., Vandaele, J., & Corthals, P. (2013). Pitch and Pitch Variation in Lesbian Women. *J Voice,* 27, 656.e13. DOI: 10.1016/j.jvoice.2013.04.008.

24. Rendall, D., Vasey, P. L., & McKenzie, J. (2008). The Queen's English: An Alternative, Biosocial Hypothesis for the Distinctive Features of "Gay Speech". *Arch Sex Behav,* 37, 188–204. DOI: 10.1007/s10508-007-9269-x.

25. Sulpizio, S., Fasoli, F., Maass, A., Paladino, M. P., Vespignani, F., Eyssel, F., & Bentler, D. (2015). The Sound of Voice: Voice-Based Categorization of Speakers' Sexual Orientation within and across Languages. *PLoS ONE*, DOI:10.1371/journal.pone.0128882

26. Kachel, S., Simpson, A. P., & Steffens, M.C. (2017b). Acoustic correlates of sexual orientation and gender-role self-concept in women's speech. Manuscript re-submitted to *J Acoust Soc Am*.

27. Kachel, S., Steffens, M. C., & Niedlich, C. (2016). Traditional masculinity and femininity: Validation of a new scale assessing gender roles. *Front Psychol 7:956.* doi: 10.3389/fpsyg.2016.00956

28. Kachel, S., Steffens, M. C., & Simpson, A. P. (2017). The Expression and Perceptions of Sexual Orientation Model: Speech based evidence. Unpublished manuscript.

29. Friedrich, Susanne (2016). Gaydar: Wer erkennt wen woran am besten? Die Rolle von Beurteilerinnenmerkmalen bei der Erkennung von sexueller Orientierung [Gaydar: Who identifies whom by which signals best? The role of rater characteristics in the detection of sexual orientation]. Unpublished Master thesis.

30. Boersma, P., & Weenink, D. (2014). PRAAT: Doing phonetics by computer (version 5.3.71) [Computer software]. http://www.praat.org/ (Last viewed September 18, 2014).

31. Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Commun*, *20*, 255–272.

32. Sandfort, T. G. M. (1997). Sampling male homosexuality. In: J. Bancroft (Ed.): Researching sexual behavior: Methodological issues. *Bloomington: Indiana University Press*, 261–275.

33. Dekker, A., & Schäfer, S. (1999). Zum Sampling-Bias empirischer Studien über männliche Homosexualität [On the sampling-bias of empirical studies about male homosexuality]. *Z Sex Forsch, 12,* 350–361.

34. Meyer, I. H., & Wilson, P. A. (2009). Sampling Lesbian, Gay, and Bisexual populations. *J Couns Psychol, 56,* 23–31. DOI: 10.1037/a0014587.

35. Ding, J. Y. C., & Rule, N. O. (2012). Gay, straight, or somewhere in between: Accuracy and Bias in the Perception of Bisexual Faces. *J Nonverbal Behav, 36,* 165–176.

36. Cox, W. T. L., Devine, P. G., Bischmann, A. A., & Hyde, J. S. (2015). Inferences About Sexual Orientation: The Roles of Stereotypes, Faces, and The Gaydar Myth. *J Sex Res, 53,* 157–171. DOI: 10.1080/00224499.2015.1015714.

Figures

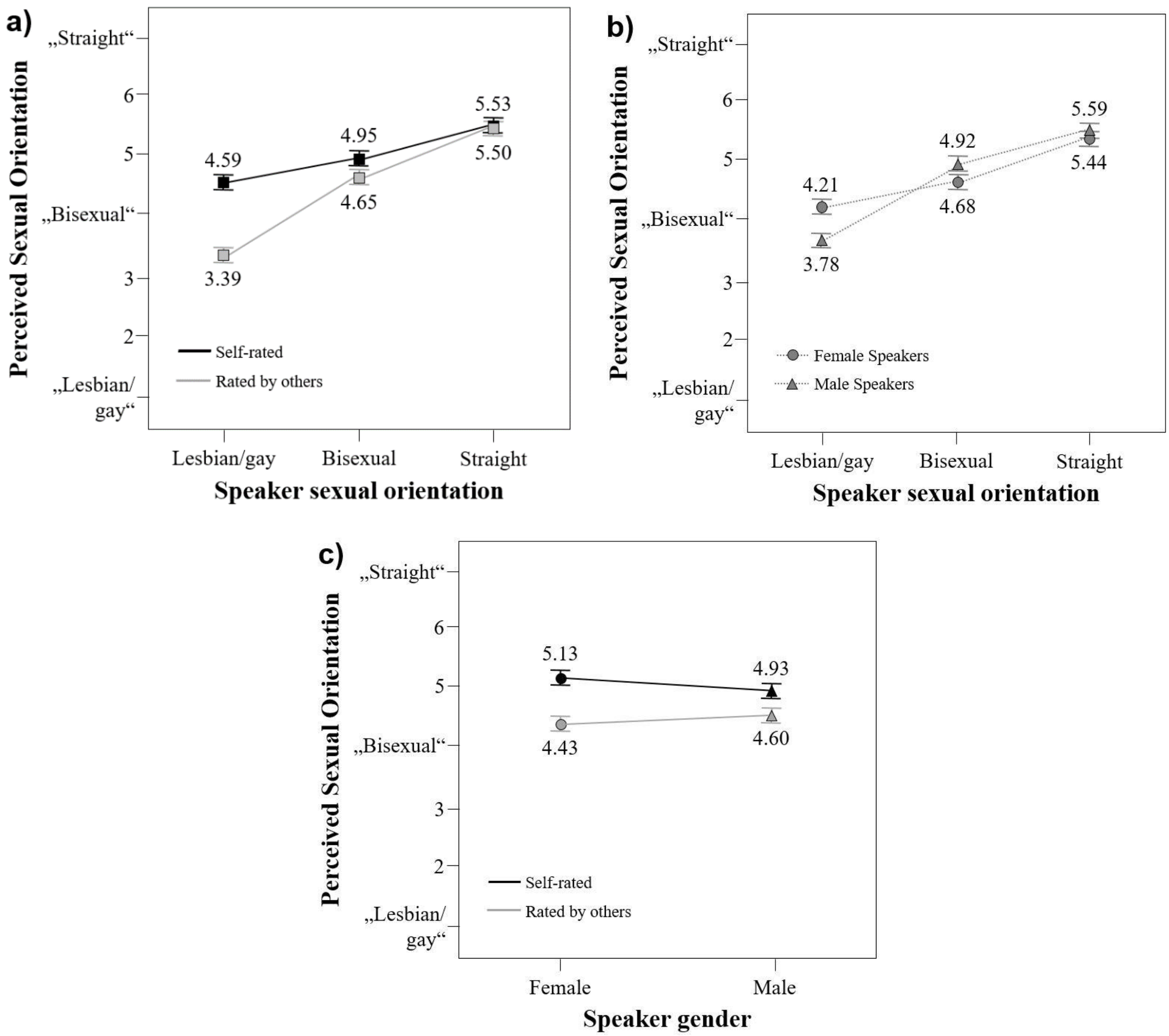


Fig 1. First-order interaction effects of speaker sexual orientation x rating basis (a), speaker sexual orientation x speaker gender (b), and speaker gender x rating basis (c) in perceived sexual orientation of voice averages.

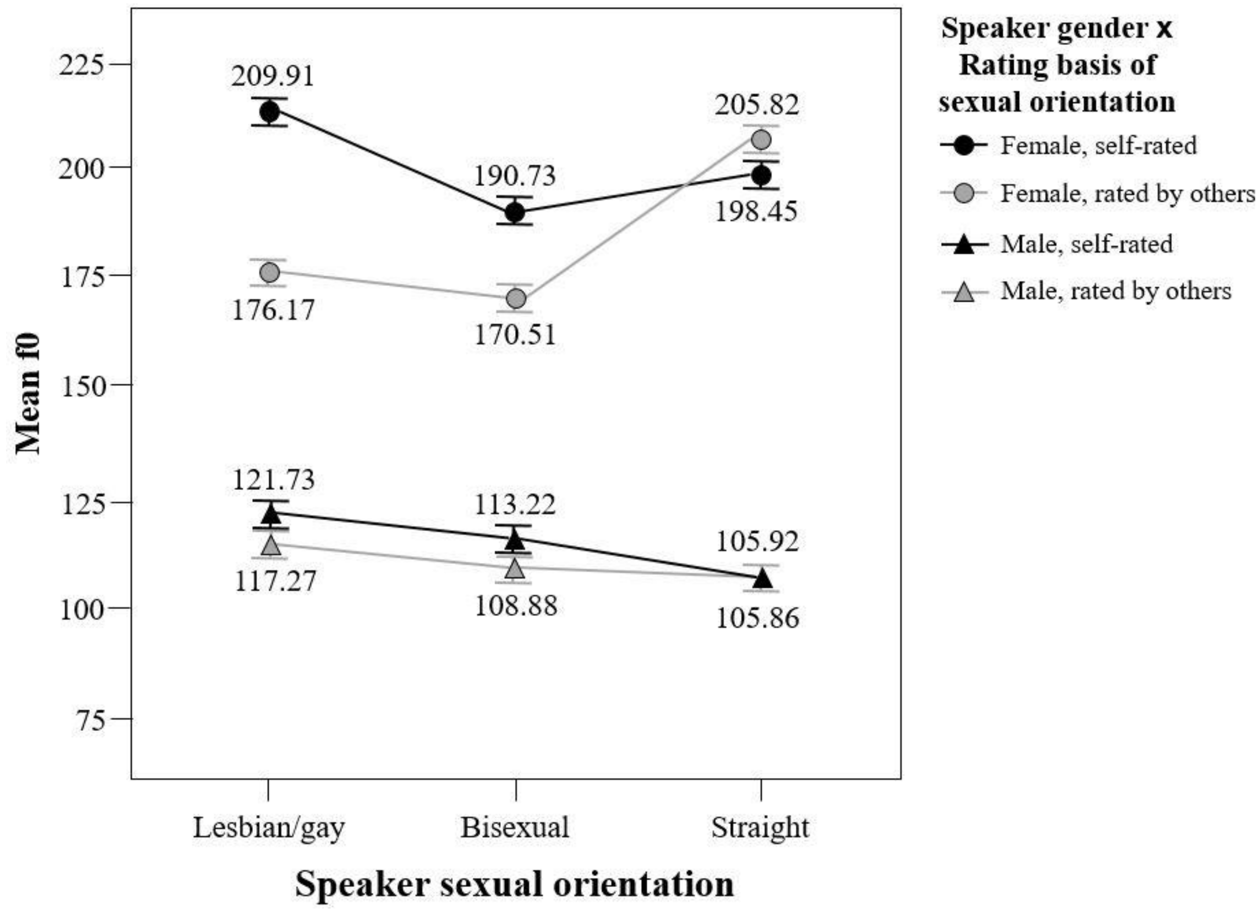


Fig 2. Mean f0 as a function of speaker sexual orientation, speaker gender, and rating basis of sexual orientation of voice averages.

Manuscript 5

Kachel, S., Steffens, M. C., & Simpson, A. P. (2017).

The Expression and Perception of Sexual Orientation Model: Speech Based Evidence.

Submitted to *Journal of Personality and Social Psychology.*

The Expression and Perception of Sexual Orientation Model: Speech Based Evidence

Sven Kachel & Melanie C. Steffens

University of Koblenz-Landau

Adrian P. Simpson

Friedrich Schiller University Jena

Key words: sexual orientation, gender-role conformity, voice, face, perception

Author Notes

Sven Kachel, Department of Social, Environmental, and Economic Psychology, Faculty of Psychology, University of Koblenz-Landau, Landau, Germany.

Please address correspondence to Sven Kachel, FB 8-Psychology, University of Koblenz-Landau, Fortstraße 7, 76 829 Landau/Pfalz, Germany. E-mail: *sven.kachel@uni-jena.de.*

Abstract

Although sexual orientation is perceptually ambiguous, people are able to detect sexual orientation of others with above-chance accuracy. In the present paper, we present the expression and perception of sexual orientation model (EPSOM) that proposes an indirect route between actual and perceived sexual orientation including three mediating components. People who differ in sexual orientation often differ in *psychological characteristics* (e.g., actual gender-role conformity) which influence different *implicit signals* (e.g., gender-related signals in voices and faces). Perceivers decode these signals and arrive at *impressions related to sexual orientation* (e.g., perceived gender-role conformity) and thus judge sexual orientation (sometimes) accurately. Predictions are derived from EPSOM and tested using speech-based evidence that provides many objective parameters for signal analysis (e.g., voice pitch). In Experiment 1, 101 raters judged the sexual orientation of 72 male and female targets differing in sexual orientations from voices, faces, and their combination. Both targets' psychological characteristics and acoustic parameters of their vocal signals were obtained. In Experiment 2, 38 raters additionally rated the gender-role conformity of voice and faces. Findings -suggest that the relationship of actual and perceived sexual orientation was mediated by the three components both separately and in combination. Most importantly, we found supporting evidence for EPSOM's overall indirect route for female targets. We discuss how future research could test additional predictions of the model.

Word Count abstract: 217 words

Imagine you received a call from an unknown number and took it. You would automatically use and integrate all information available in order to classify the person on the other end of the telephone line. Similarly, you do so when you look at facial photographs of others in social network media. Research on social perception has repeatedly shown that gender, age, and regional origin were judged with a high degree of accuracy. However, detection rates of perceptually ambiguous social categories such as sexual orientation have also been typically above chance, albeit far from perfect (for a review, see Tskhay & Rule, 2013a). Previous research has repeatedly established that actual sexual orientation is related to perceived sexual orientation even given impoverished stimuli such as single words (Munson, McDonald, DeBoe, & White, 2006) or cropped faces (Rule, Ambady, Adams, & Macrae, 2008). What is missing is a theoretical account specifying how, under which conditions, and for/by whom sexual orientation is perceived correctly.

The aim of the present research is to introduce a comprehensive model that details how people express their actual sexual orientation and how observers decode sexual orientation when perceiving it. This model predicts that targets who differ in their sexual orientation also differ in a range of psychological characteristics, that these characteristics are conveyed via speech-related and facial signals, and that these signals are used for forming impressions related to sexual orientation, which in turn lead to detecting sexual orientation. Hence, we investigate raters' accuracy of correctly identifying sexual orientation (Experiment 1) and gender-role conformity categorization (Experiment 2) of female and male targets who differ in sexual orientation and other psychological characteristics when presenting different cues as proposed by the model: auditory cues (voices), visual cues (faces), and audio-visual cues (combining faces and voices).

The Expression and Perception of Sexual Orientation Model (EPSOM)

Inspired by a standard communication model (Shannon, 1948), we suggest that there is both a direct and an indirect route by which sexual orientation can be expressed, transmitted, and decoded. Both routes depend on the situation (e.g., lesbians and gay men possibly try to hide their sexual orientation when they are confronted with a homonegative environment). We call this model the Expression and Perception of Sexual Orientation Model (EPSOM, see Figure 1). Different from prominent perception models (e.g., Face-Processing Model by Bruce & Young, 1986; or Dynamic Interactive Theory of Person Construal by Freeman & Ambady, 2011), EPSOM focuses specifically on sexual orientation. Moreover, EPSOM goes beyond a detailed analysis of low- or higher-level perceptual mechanisms and thus helps us to understand how actual and perceived sexual orientation are connected to each other.

On both routes proposed by EPSOM, signals function as mediators. *Explicit signals* of actual sexual orientation established in a given cultural context form the direct route. If perceivers are aware that a rainbow flag signals a café owner's sexual orientation, the amazons' double-ax necklace is used to symbolize lesbianism, and a photo of a woman in a wedding dress with a man in a tuxedo by her side signals straightness, they will correctly identify the socially relevant symbols and by extension the actual sexual orientation of the person displaying the symbol will be accurately perceived. The only precondition is cultural knowledge pertaining to these signals that may differ between individuals (e.g., a red ribbon may function as a signal for some, but not all perceivers). Sexual orientations beyond lesbianism, gayness, and straightness are less rooted in cultures (e.g., missing bars for bisexual and pansexual women and men). Hence, EPSOM deals with the sexual orientations of lesbians, gay men, straight women and men (for previous research on bisexuals see Ding & Rule, 2012).

Expression and Perception of Sexual Orientation Model

The more elaborated aspect of EPSOM is the indirect route which consists of five components (i.e., actual sexual orientation, psychological target characteristics, implicit signals, impressions related to sexual orientation, and perceived sexual orientation) that are connected by paths (e.g., path between actual sexual orientation and psychological target characteristics). We argue that in the absence of explicit signals it is not actual sexual orientation that can be perceived by others. Instead, target's actual sexual orientation is related to a range of psychological characteristics (e.g., actual gender-role conformity). These target characteristics, in turn, can be transmitted by *implicit signals* (e.g., voices, faces). Picking up on those signals, perceivers arrive at impressions of targets (e.g., perceived gender-role conformity) and may conclude from those impressions what an individual's sexual orientation is. With the aid of the three mediating components, EPSOM is able to explain inaccuracies in the perception of sexual orientation. For example, a lesbian who self-describes as rather feminine produces gender conforming speech patterns leading to a perception as feminine, and hence may be assumed to be straight. By integrating perspectives of targets and perceivers, EPSOM can account for divergent findings of previous studies that largely disregard psychological characteristics of targets that we consider to be mediating components. For example, a gender non-conform straight and a gender-conform gay target sample possibly explains why straight men were rated as more feminine than gay men in one study (Valentova & Havlíček, 2013).

Whereas evidence of many of the single paths specified in EPSOM exists in the literature, the model as a whole contributes to the research question how ambiguous group membership based on sexual orientation can (sometimes) be perceived. In the following three sections, we describe the proposed indirect route of detecting sexual orientation by focusing on the three mediating components EPSOM postulates and by reviewing the evidence of their single paths.

Implicit Signals of Actual and Perceived Sexual Orientation

EPSOM posits that different implicit signals can be used for expressing and perceiving sexual orientation. What implicit signals and which of their characteristics have been shown to transmit sexual orientation information? After some general information on facial and vocal signals (for research on body shape and movement see Johnson, Gill, Reichman, & Tassinary, 2007; Ambady, Hallahan, & Conner, 1999), we ask which signal characteristics are correlates of actual and perceived sexual orientation.

Comparison of Faces and Voices

People's ability of correctly perceiving others' sexual orientation by using implicit signals (Shelp, 2002) is often called gaydar (Woolery, 2007). Gaydar studies have almost consistently shown that actual sexual orientation can be correctly perceived. Sexual orientation of lesbians, gay men, straight women, and straight men was perceived with above-chance accuracy based on *faces* (Rule & Ambady, 2008, Rule et al., 2008; Rule, Ambady, & Hallett, 2009; Hughes & Bremme, 2011; Rule, 2011; Rule, Ishii, Ambady, Rosen, & Hallett, 2011; Rule, Rosen, Slepian, & Ambady, 2011; Valentova & Havlíček, 2013; Stern, West, Jost, & Rule, 2013; Tabak & Zayas, 2013; Tskhay, Feriozzo, & Rule, 2013; Lick & Johnson, 2014a; Tskhay & Rule, 2015; Tskhay, Krendl, & Rule, 2016) and *voices* (Kachel, Simpson, & Steffens, 2017a; Gaudio, 1994; Linville, 1998; Smyth, Jacobs, & Rogers, 2003; Munson et al., 2006; Rieger, Linsenmeier, Gygax, Garcia, & Bailey, 2010; Zimman, 2010; Valentova & Havlíček, 2013; van Borsel & van de Putte, 2014; Sulpizio, Fasoli, Maass, Paladino, Vespignani, Eyssel, & Bentler, 2015: Study 2; Renn, 2003). Only a small number of voice-based studies failed to replicate this finding (Sisson, 2003; Piccolo, 2008; Sulpizio et al., 2015: Study 1). Voice-based studies used a variety of stimuli ranging from small read groups of words (Munson et al., 2006) to longer excerpts of spontaneous speech (Smyth et al., 2003; Rieger et al., 2010). Few studies have investigated female targets (Munson et al., 2006; Piccolo, 2008; Rieger et al., 2010). For face-based studies, male targets are also overrepresented. Moreover, face-based studies almost entirely collected target pictures from U.S. online-dating platforms. Hence, quality differences of lesbian/gay vs. straight facial photographs could explain good gaydar performance (Cox, Devine, Bischmann, & Hyde, 2006). However, there are two less ambiguous findings. First, when pictures were taken in a lab under standardized conditions gay men still correctly rated men’s sexual orientation (Valentova & Havlíček, 2013). Second, correct classifications were still obtained when photos were posted by targets’ acquaintances (low self-presentation bias; Rule & Ambady, 2008; Rule et al., 2008). Hence, research is missing that compares female and male target samples, and more evidence is needed on non-U.S. targets and using photos where picture-inherent quality differences are ruled out.

Although there is little research on the relative importance of voices and faces in the categorization of sexual orientation (Sulpizio et al., 2015), it has been hypothesized that voices should lead to better judgments than visual information (see Tskhay & Rule, 2013a) because the auditory channel is particularly revealing. Indeed, voice recordings led to more correct categorizations of male targets than facial pictures; however, voice-based and face-based perceptions did not correlate (Valentova & Havlíček, 2013). Moreover, some studies suggest that additional information elicits higher accuracy. Correct sexual orientation categorizations significantly increased from presenting a photo series over a short video clip devoid of speech information to longer videos (Ambady et al., 1999). However, there seems to be an influence of targets’ gender: Whereas for male targets, videos with sound led to more correct judgments than speech which was in turn a better signal than full appearance, for female targets, sexual orientation information contained in videos did not exceed full appearances but both exceeded speech (Rieger et al., 2010). Hence, research is missing that further compares different implicit signals (voices vs. faces vs. combined voice+face stimuli).

Path from Actual Sexual Orientation to Implicit Signals

Only a small number of studies investigated facial structures and asymmetry as correlates of actual and perceived sexual orientation (Hughes & Bremme, 2011; Valentova, Kleisner, Havlíček, & Neustupa, 2014; Skorska et al., 2014). Faces of straight women and men showed less asymmetry (Hughes & Bremme, 2011; but see Valentova et al., 2014 for male faces) and more gender conforming facial structures than lesbians' and gay men's faces (Skorska, Geniole, Vrysen, McCormick, & Bogaert, 2014; but see Hughes & Bremme, 2011). Hence, there is some evidence that targets' face structures mirror their actual sexual orientation.

Is actual sexual orientation reflected in targets' voices? Interindividual differences in acoustic parameters do not depend on biophysical inevitabilities only, but are additionally determined by learned behavior (Simpson, 2009). A first class of acoustic parameters commonly used in the sociophonetic studies on sexual orientation are intonational features that encompass a variety of characteristics involving fundamental frequency – frequency of vocal fold oscillation per second – which is the acoustic correlate of voice pitch. These characteristics mainly describe the fundamental frequency distribution of a speaker's utterance such as mean fundamental frequency (corresponding to mean voice pitch), maximum f0 (corresponding to upper voice pitch boundary), and fundamental frequency range (corresponding to the difference between upper and lower voice pitch boundaries).

Because German male speakers produce lower mean fundamental frequencies (100-120 Hz) than German female speakers (200-220 Hz), they are perceived as sounding lower pitched (Simpson, 2009). Regarding sexual orientation, findings on intonational features are inconclusive for female (for a review see Kachel, Simpson, & Steffens, 2017b) and male speakers (for a review see Kachel et al., 2017a). Corroborating stereotypes, some studies found that straight women use a higher mean fundamental frequency than lesbians (Camp, 2009; van Borsel, Vandaele, & Corthals, 2013) and straight men use a lower mean f0 than gay men (Linville, 1998; Baeck, Corthals, & van Borsel, 2011). Straight women also showed a higher maximum fundamental frequency, but the fundamental frequency range among lesbians was more varied than that among straight women (Camp, 2009), suggesting that it may depend on the specific sample recruited whether lesbians on average speak with a lower pitch than straight women. In line with this idea, other studies failed to find a significant effect of sexual orientation on mean fundamental frequency in women (Kachel et al., 2017b; Munson et al., 2006; Rendall, Vasey, & McKenzie, 2008) and men (Kachel et al., 2017a: Lerman & Damsté, 1969; Gaudio, 1994; Sisson, 2003; Munson et al., 2006; Rendall et al., 2008; Zimman, 2010; Valentova & Havlíček, 2013; Sulpizio et al., 2015).

A second class commonly used in sociophonetic studies on sexual orientation is vowel space characteristics. Different vowel qualities are the result of different vocal tract configurations and are typically described using two measures: the first and the second formant frequency. Within the German vowel system, /a/ shows the highest first formant frequency (henceforth F1) because it is linked to the highest distance between the tongue dorsum and the hard palate, /i/ shows high second formant frequency (henceforth F2) because the tongue is positioned in the very front of the mouth and is produced with spread lips, whereas /u/ is produced with rounded lips and the tongue in the back of the mouth and exhibiting low F2 (see Figure A1 in the Online Appendix). The vowel space – the polygon formed by the distances between the single vowels in the two-dimensional acoustic space – is on average more expanded and more shifted to the front and down for female than male speakers. Regarding sexual orientation, findings for vowel space characteristics are similarly ambiguous as for intonational features. While some studies found vowel space shift (Pierrehumbert, Bent, Munson, Bradlow, & Bailey, 2004; Rendall et al., 2008) and vowel space expansion (Rendall et al., 2008) to be in line with stereotypes for women, others did not (Munson et al., 2006; Kachel et al., 2017b). The same is true for male speakers: Straight men were shown to have acoustically more expanded vowel spaces than gay men in one study (Pierrehumbert et al., 2004). However, this finding could not be replicated (Munson et al., 2006; Rendall et al., 2008; Kachel et al., 2017a). The evidence for single vowel formants is even more inconsistent (for a review see Kachel et al., 2017a, b). Hence, only some studies showed acoustic correlates of actual sexual orientation.

Path from Implicit Signals to Perceived Sexual Orientation

Do signal characteristics influence others' perceptions of targets' sexual orientation? To begin with facial features, men but not women were judged as straighter, the less facial asymmetry they had (Hughes & Bremme, 2011). However, raters did not use any of the gender-conform facial structures to determine female and male targets' sexual orientation (Hughes & Bremme, 2011; Valentova et al., 2014). Using a more indirect approach, raters were presented with different facial details. Hairstyle and eyes were shown to carry information about women's and men's sexual orientation (Rule et al., 2008; Rule, Ambady, et al., 2009; Tskhay et al., 2013; but see Cox et al., 2016). However, none of the facial details is itself necessary for correct gaydar performance (Rule et al., 2008) which can be explained in terms of configural face processing (Tabak & Zayas, 2012). When facial structures provided contradicting gender information (e.g., feminine face shape and masculine texture), the faces were judged as less straight (Freeman, Johnson, Ambady, & Rule, 2010). Thus, there are facial correlates of actual and perceived sexual orientation. However, facial information that raters used to determine targets' sexual orientation seem to differ from those facial features reflecting targets' actual sexual orientation.

Regarding the question as to which voice features are used for sexual orientation judgments, results were again inconclusive (for a review focusing on men, see Kachel et al., 2017a). The finding that female voices were judged as straighter the lower their mean pitch and less variable their pitch (Camp, 2009) was not replicated (Munson et al., 2006). Although women were perceived as straighter the higher their mean F1 and F2, findings on vowel space expansion were inconsistent (Munson et al., 2006; Piccolo, 2008). A possible reason for inconsistent findings regarding intonational and vowel space characteristics is that studies of the phonetic correlates of sexual orientation have largely failed to take into account another mediator (see Figure 1): the psychological characteristics that possibly explain within-group variance (see also Waksler, 2001; Munson & Babel, 2007).

Targets' Psychological Characteristics

Path from Actual Sexual Orientation to Psychological Characteristics

Decades of research testing psychological differences between straight women and men on the one hand and lesbians and gay men on the other have established a whole range of psychological characteristics related to sexual orientation; among others, there are differences in the inter-related factors actual gender-role conformity, social group affiliation, and social environment. Many studies have reported that straight women and men showed higher actual gender-role conformity than lesbians and gay men (e.g., Lippa, 2005; Spence & Helmreich, 1975; Kachel, Steffens, & Niedlich, 2016; Rieger, Linsenmeier, Gygax, & Bailey, 2008). Still, there is much variance in actual gender-role conformity within straight women and lesbians (Kachel et al., 2016). The latter described themselves as either "butch" (i.e., more masculine) or "femme" (i.e., more feminine; Loulan, 1990; Pearcey, Docherty, & Dabbs jr, 1996; also see Singh, Vidaurri, Zambarano, & Dabbs, 1999). Thus, it depends on the individual to what degree actual straightness is related to actual gender-role conformity.

The social-identity perspective postulates that people define themselves not only as individuals, but also as members of social groups that they identify with, in turn self-stereotyping to align with the respective social-group stereotype (Tajfel & Turner, 1979; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). For instance, some lesbians were less likely to wear make-up and often shortened their hair after they came out (Krakauer & Rose, 2002). Thus, to the degree that their sexual orientation is a valued and important part of their *social group affiliation*, straight women and men as well as gay men and lesbians should construe their self-concepts in line with the respective social-group prototypes (i.e., the typical member of the group in a given social context).

Further differences in psychological characteristics that depend on sexual orientation could be related to different *social environments*. To the degree that lesbians, gay men, and straight women and men are exposed to different role models, different new and subgroup-specific implicit signals may be created to signal sexual orientation (see Smyth & Rogers, 2008; Pierrehumbert et al., 2004, for similar ideas). For example, lesbians whose friendship networks consist mainly of lesbians could adopt the habit to express their sexual orientation more than lesbians whose friendship networks consist mainly of straight women. Taken together, there is a lot of evidence that people's actual sexual orientation and their psychological characteristics are related (i.e., the first path in EPSOM, see Figure 1).

Path from Psychological Characteristics to Implicit Signals

To the best of our knowledge, no face-related and only a few voice-related studies in the context of sexual orientation have tested which targets' psychological characteristics influence the relationship between actual sexual orientation and signals. Although the first socio-phonetic study found no correlation between gay men's mean fundamental frequency and their coming-out age and extent (Baeck et al., 2011), Kachel and colleagues (2017a, b) shed light on mediating variables. For instance, they showed that the more masculine male speakers described themselves, the lower their mean F2 (also see Biemans, 2000). Additionally, they found lesbians' and gay men's speech parameters to be related to several psychological characteristics (e.g., actual gender-role conformity, social environment). EPSOM accounts for research on implicit signals reflecting psychological characteristics (i.e., the second path in EPSOM, see Figure 1).

Path from Psychological Characteristics to Perceived Sexual Orientation

Which target characteristics lead to a better perception of targets' sexual orientation? Whether perceivers are more accurate when judging women or men depends on the signal. Given videos depicting the whole person (Ambady et al., 1999; Sylva, Rieger, Linsenmeier, & Bailey, 2010) or photos of faces (Tabak & Zayas, 2012; Lyons, Lynch, Brewer, & Bruno, 2014), women were judged more accurately than men, but given videos of silhouettes, men were judged more accurately than women (Ambady et al., 1999; Johnson et al., 2007). However, there is no voice-related study testing whether perceivers are more correct in judging women's or men's sexual orientation.

Whereas some studies focused on gender and other sociodemographic target characteristics such as ethnicity/race (e.g., Johnson & Ghavami, 2011; Rule, 2011), nationality/language (e.g., Sulpizio et al., 2015), and age (e.g., Tskhay et al., 2016), only a few studies dealt with psychological target characteristics. More concealment motivation tended to be associated with being perceived as straighter for gay but not lesbian targets (Sylva et al., 2010). Moreover, the less positive attitudes gay speakers held regarding sounding gay (Mann, 2012), the straighter they were perceived. Thus, although there are some pieces of evidence that psychological target characteristics are important for perceiving targets' sexual orientation, their importance needs to be tested systematically.

Impressions Related to Perceived Sexual Orientation

There is a wealth of research on impressions regarding sexual orientation related traits, like affect (Lyons et al., 2014; Tskhay et al., 2016), attractiveness (Lyons et al., 2014), and speech clarity (Munson et al., 2006). All of these traits are gendered (e.g., positive affect is stereotypically female). According to lay people's gender inversion theory (Kite & Deaux, 1987), the central aspect of stereotypes of lesbians is that they are more masculine (and less feminine) than straight women, whereas gay men are assumed to be more feminine and less masculine than straight men (e.g., Kite & Whitley, 1996). Similarly, gender inversion is expected as a cue to sexual orientation: Straight men who showed stereotypical feminine behavior expected being misidentified as gay more often than straight men who showed stereotypical masculine behavior (Bosson, Prewitt-Freilino, & Taylor, 2005). In line with the "kernel of truth" hypothesis of stereotyping (Prothro & Melikian, 1955), perceived gender-role conformity differences correspond to actual differences reported above. Hence, we focus on perceived gender-role conformity as the primary and encompassing impression related to perceived sexual orientation.

Path from Actual Sexual Orientation to Impressions Related to Sexual Orientation

Gender-role conformity judgments of raters depend on targets' actual sexual orientation given a variety of signals. Straight women and men were perceived as more gender conforming (and less variable in their gender-role conformity, Rieger et al., 2008, Rieger et al., 2010) than their lesbian/gay counterparts based on their full appearance shown in photos (Rieger et al., 2010) and videos (Rieger et al., 2008; Rieger et al., 2010; Valentova, Rieger, Havlíček, Linsenmeier, & Bailey, 2011), based on videos of point-light ball displays showing targets' moving contours (Lick & Johnson, 2013), speech recordings (Gaudio, 1994; Munson, 2007; Rieger et al., 2010), and even handwritings (Lester, McLaughlin, Cohen, & Dunn, 1973). However, mixed results were found for men's faces. The finding that straight men were judged to be more masculine/less feminine (Stern et al., 2013; Lyons et al., 2014) was not replicated (Tskhay & Rule, 2015). Additionally, other target (e.g., age; see Tskhay et al., 2016) and stimulus characteristics (e.g., phonological structure of words; see Munson, 2007) influenced the association of actual sexual orientation and perceived gender-role conformity. EPSOM integrates the influences of both components. Moreover, it allows for a more complex picture by accounting for the finding that actual sexual orientation was better predicted by perceived gender-role conformity than perceived sexual orientation (Valentova et al., 2014) which suggests a mediation by impressions related to sexual orientation.

Path from Psychological Characteristics to Impressions Related to Sexual Orientation

Which relations do exist between impressions related to sexual orientation, especially perceived gender-role conformity, and psychological target characteristics? After ruling out possible memory biases, lesbian/gay and straight people who were more gender conforming during childhood became more gender conforming adults and were in turn perceived as more gender conforming by others based on videos (Rieger et al., 2008). Besides actual gender-role conformity in a narrow sense, psychological characteristics related to it such as sexual role were connected to perceived gender-role conformity: Gay men who self-described as tops (i.e., insertive partner during anal intercourse) were perceived as more masculine than bottoms based on facial photographs (i.e., receptive partner during anal intercourse; Tskhay & Rule, 2013b). Regarding acoustic signals, no study so far investigated whether speech recordings were able to transmit gender-role conformity. Hence, there is a research gap on whether actual and perceived gender-role conformity are linked based on voices.

Path from Implicit Signals to Impressions Related to Sexual Orientation

Although the visual domain is represented very well in gaydar studies, we found only a few studies on the signaling of gender-role conformity for this area. Wider faces, massive jaws, rounded chins, thinner lips, smaller eyes, and thicker eyebrows (Valentova et al., 2014) as well as a tube-shaped body and a broad-shouldered gait were perceived as masculine (Johnson et al., 2007). Most research on this topic was done for the acoustic field. Disregarding anecdotal evidence (n = 2; Guzik, 2004), there are inconsistent result patterns for intonational and vowel space correlates of perceived gender-role conformity. Women displaying high-pitched voices (Camp, 2009; Munson, 2007) and more variable fundamental frequencies (Camp, 2009) were rated as more feminine; however, the finding that men displaying low-pitched voices (Sulpizio et al., 2015: German men; Terango, 1966; Munson, 2007; Valentova & Havlíček, 2013) and less variable fundamental frequencies (Terango, 1966) were rated as more masculine, was not replicated in some studies (Camp, 2009; Sulpizio et al., 2015: Italian men; Smyth et al., 2003). The pattern is similarly inconclusive for vowel space characteristics. Whereas vowel space dispersion and mean F2 were acoustic correlates for perceived gender-role conformity in women (Munson, 2007), mixed results turned out for men (Munson, 2007; Sulpizio et al., 2015). EPSOM accounts for this evidence on associations between implicit signals and impressions related to sexual orientation by suggesting the third path of its indirect route.

Path from Impressions Related to Sexual Orientation to Perceived Sexual Orientation

Can impressions of gender-role conformity be used to deduce targets' sexual orientation? Perceived straightness and perceived gender-role conformity were correlated with each other independent of the signal (Rieger et al., 2010). Female and male targets who were rated as straighter were also judged as more gender conforming based on faces (Dunkle & Francis, 1990; Freeman et al., 2010; Valentova & Havlíček, 2013; Valentova et al., 2014; Tskhay & Rule, 2015, Tskhay et al., 2016), voices (Gaudio, 1994; Smyth et al., 2003; Camp, 2009; Rieger et al., 2010; Sulpizio et al., 2015), videos of full appearances (Valentova et al., 2011; Rieger et al., 2010), and light-point displays of targets' moving contours (Lick & Johnson, 2013). However, no cross-modal correlations were found (Valentova & Havlíček, 2013); for example, perceived straightness judged from faces was not connected to perceived gender-role conformity based on voices. The correlation of perceived straightness and perceived gender-role conformity depended on targets' gender: Compared to male targets, judgements about women's sexual orientation were inferred to a higher degree from gender-role conformity perceptions (Munson, 2007; Lick & Johnson, 2014b). Moreover, there are different male voice features indicating perceived straightness and perceived gender-role conformity (Munson, 2007; Kachel et al., 2017a) and there are low male voices that sound gay but not feminine whereas sounding gay and feminine is almost the same for high male voices (Smyth et al., 2003). Hence, although perceived gender-role conformity and perceived sexual orientation are not the same, perceived gender-role conformity is a cue for judgments about female and male targets' sexual orientation. Consequently, EPSOM proposes that judgments about sexual orientation base on other perceptions, primarily gender-role conformity impressions, that are decoded from voices, faces, and other signals (see Munson et al., 2006, and Freeman et al., 2010, for similar ideas). Hence, EPSOM accounts for this evidence by forming the fourth path of its indirect route.

The Present Research

As delineated above, EPSOM integrates much evidence by postulating three components mediating the effect of actual on perceived sexual orientation. The present research addresses two batteries of questions. The first battery regards open questions referring to EPSOM's single mediating components:

- Regarding the "Implicit Signals" component we ask: Do raters perceive sexual orientation from non-U.S. target facial photos with higher than chance accuracy when picture-inherent quality differences are ruled out? Does the combination of voices and faces lead to higher accuracy than the single signals? Do raters perceive sexual orientation from female vs. male target voices more accurately? Could judgments based on one signal be used to deduce judgments based on another signal (i.e., is a person who is seen as straight, heard as straight)? Is voice a more valid signal for actual sexual orientation and gender-role conformity compared to face? Which acoustic parameters are connected to perceived sexual orientation and gender-role conformity in women and men?
- Regarding the "Psychological Characteristics" component we wanted to know: Which psychological target characteristics are connected to perceived sexual orientation and gender-role conformity?
- Regarding the "Impressions related to sexual orientation" we are interested in: Are actual and perceived gender-role conformity based on voices linked to each other?

The second battery of questions regards EPSOM's indirect route mediating the effect of actual on perceived sexual orientation:

- Do the single EPSOM components or their combinations mediate the relationship between actual and perceived sexual orientation?
- Is perceived gender-role conformity an additional mediator (in combination with actual gender-role conformity and acoustic parameters) of the relationship between actual and perceived sexual orientation?

In the present study, we compare female and male targets, use non-U.S. samples, and contrast different signals, namely voices, faces, and their combination. Other than previous studies, we use face stimuli taken under standardized lab conditions in order to control for quality differences between lesbian/gay and straight targets (see Cox et al., 2015). Because of the predominance of voice-based research in providing evidence for specific signal characteristics, we also focused on the most common acoustic parameters (i.e. intonational and vowel space features). In order to provide explanations for inconsistent acoustic and facial correlates of actual and perceived sexual orientation in previous studies, we collected data on psychological characteristics (i.e., actual gender-role conformity, social group affiliation, and social environments). In Experiment 1, we investigate perceived sexual orientation, in Experiment 2 we investigate perceived gender-role conformity.

Experiment 1

The main aim of Experiment 1 was to provide first evidence for EPSOM's indirect route and to answer open questions on its mediating components. We collected data on perceived sexual orientation based on three different signals of a sexually divergent women's and men's target sample. Regarding EPSOM's "Implicit Signals" component, we investigated if sexual orientation based on voice signals could be better perceived compared to facial presentations, if an improvement of correct judgments could be made by combining both signals and if there are gender differences, as suggested by previous findings. Moreover, we were interested in whether people who are heard as straight are also seen as straight and if there are acoustic correlates of perceived sexual orientation. Regarding EPSOM's "Psychological Characteristics"-component, we wanted to find out which psychological features (e.g., actual gender-role conformity) played a prominent role in raters' sexual orientation judgments.

Method

Participants

Targets

We recruited 111 volunteers in a German metropolis with more than 1,000,000 inhabitants (female participants only; *reference deleted for anonymous review*) and a German city with approximately 100,000 inhabitants (male participants only, *reference deleted for anonymous review*) who were invited to participate in an investigation on "Sexuality and Voice" with different sexual orientations. Targets were recruited in a university context and through LGBTIQ* (Lesbian, gay, bisexual, transgender, intersexual, and queer) organizations and platforms. A precondition for participation was that people were native German speakers without voice and speaking disorders. The sample was highly homogeneous with regard to ethnicity, educational level, and age: nearly all participants held a university entrance degree, were Caucasian, and ranged in age from 20 to 30 years.

97 gave permission for the use of either voices and photographs. After excluding the few bisexual targets, we randomly selected 18 lesbians, 18 gay men (self-ratings on a Kinsey-like 7-point scale: 1-2), 18 straight women, and 18 straight men (Kinsey-like scores: 6-7)

from this sample. Mean age was 24.2 years (*SD* = 2.26). The four target groups did not differ in any sociodemographic variable.

Raters

We recruited 101 raters (65 women and 36 men). Most of them (*n* = 81) were students at a mid-sized German university. The others were passers-by recruited in the city center asking for participation in a study that tested if it was possible to correctly judge a person's sexual orientation by either hearing their voice, seeing their face, or both. Age ranged from 18 to 71 years, average age was 28.00 (*SD* = 12.84).

Materials

Stimuli

Auditory and visual recordings from our 72 selected targets were used. Six additional stimuli were used as practice stimuli (three female and three male, independent of their sexual orientation). For all selected targets a digitalized photograph of the face and a digitalized voice recording were available.

Photographs

All photographs were in color and showed a smiling face, looking directly at the camera. The photographs were cropped so that only the face with and the neck as well as a bit of the shoulder were visible in front of the a light background. The targets were also allowed to have any amount of jewelry, make-up, facial hair etc., which resulted in face stimuli with a high ecological validity.

Voice Recordings

As voice stimuli we used recordings of the sentence <Der Tag ist sehr lang geworden>

(/deːɐ̯ tʰaːk ɪst z̥eːɐ̯ laŋ ɡəvɔʁdn̩/; English: “It has been quite a long day.”). This sentence was chosen out of a list of 20 sentences because it contains many phonetic aspects discussed in previous research on the perception of sexual orientation (e.g., /aː/, /s/, /ɪ/, see Kachel et al., 2017b).

Acoustic Measures

In order to determine which acoustic features accompany a certain sexual orientation rating, we acoustically analyzed the recordings of all targets. The stimulus sentence served as a basis for acoustic measurement for fundamental frequency features and vowel space characteristics. Different features of the individual fundamental frequency distribution were measured by means of the whole sentence: mean (indicating average fundamental frequency), standard deviation (indicating fundamental frequency variability), 2.5^{th} percentile (indicating lower fundamental frequency boundary), 97.5^{th} percentile (indicating upper fundamental frequency boundary). Vowel space characteristics were determined on the basis of three vowels (/aː/ in <Tag>, /ɪ/ in <ist>, /ɔ/ in <geworden>). For each vowel first and second formant frequencies were measured using the formant tracking function in the software program *praat* (Boersma & Weenink, 2014). According to Pierrehumbert et al. (2004) we computed vowel space expansion using the Bradlow, Torretta, and Pisoni (1996) method which is measured as the mean Euclidian distance of the three vowels from the center of the vowel triangle and vowel space shift which is indicated by mean F1 (degree to which the vowel space is directed to the bottom of the mouth) and mean F2 (degree to which the vowel space is directed to the front of the mouth) across the three vowels.

Psychological Measures

A whole range of psychological characteristics were collected from the targets that we classified into four groups. For the purpose of easier interpretation all psychological measures were recoded such that high scores indicate straighter and/or more gender conforming characteristics.

Straight Attraction Index

We collected data on sexual orientation with different measures. First, we used a Kinsey-like scale ("Regarding my sexual orientation I self-identify as...") ranging from 1 (exclusively gay/lesbian) to 7 (exclusively straight). Second, we used more objectifiable measures for sexual orientation. The same four items (sexual fantasy, romantic feelings, physical attraction, and sexual interaction) were rated separately regarding women and men. Targets were asked to specify their sexually related sensations and experiences by means of frequencies on a 7-point scale ranging from 1 (*never*) to 7 (*always*). Two gender-separated means were calculated separately across the four corresponding items in order to determine the degree of sexual orientation towards women and men. All means were re-coded in order to associate high scores with a high degree of straight attraction. Because of high inter-correlations of subjective and more objectifiable measures (all $r > .89$, all $p < .001$), a straight attraction index was created by computing the mean of the Kinsey-like scale and straightness conform orientation towards men and women ($\alpha = .98$).

Actual Gender-Role Conformity

Several instruments were used to determine gender-role conformity. Data of present, personality- and behavior-based gender-role conformity were collected as well as data of childhood gender-role conformity.

Traditional Masculinity-Femininity Scale. The first instrument we applied was the Traditional Masculinity-Femininity scale (TMF; Kachel et al., 2016) that consists of six items measuring different aspects of actual gender-role conformity according to Constantinople (1973) such as gender-role adoption ("I consider myself as..."), gender-role preference ("Ideally, I would like to be..."), and gender-role identity ("Traditionally, my interests/attitudes and beliefs/behavior/outer appearance would be regarded as..."). Targets were asked to rate each item on a 7-point scale ranging from 1 (*totally masculine*) to 7 (*totally feminine*). Internal consistency was α = .91.

Femininity-Scale of the German Extended Personal Attributes Questionnaire. We used the expressiveness scale (GEPAQ-F) of the German Extended Personal Attributes Questionnaire (Runge, Frey, Gollwitzer, Helmreich and Spence 1981), which is the German version of the Extended Personal Attributes Questionnaire by Spence and Helmreich (1975). We did not include the instrumentality scale (GEPAQ-M), because women and men converged in instrumental traits in recent years, while expressive traits still differed (Kachel et al., 2016). Comprising eight items, the GEPAQ-F measures personality traits more socially desirable for women than men (e.g., emotional). Targets rated themselves regarding each item on a 5-point scale ranging from 1 (*not emotional*) to 5 (*very emotional*, α = .71).

Feminine Gender-Role Conforming Behavior. The scale consists of 29 everyday behaviors more typical for women than men (Athenstaedt, 2003). Targets rated themselves regarding typically feminine behaviors (e.g.,"put flowers on the desk") on a 6-point scale ranging from 1 (*completely atypical for me*) to 6 (*completely typical for me*; α = .92).

Childhood Gender-Role Conformity. 9 Items, such as "I played typical girls' games." or "I behaved like a typical girl" (examples were taken from the women's version), were applied in order to assess Childhood Gender-Role Conformity (CGRC, see Kachel et al., 2016). Targets self-described their typicality on a 5-point scale ranging from 1 (*I strongly disagree*) to 5 (*I strongly agree*; $\alpha = .87$).

Social Group Affiliation

Two 7-point circle scales were used to assess the affiliation of the self with the group of same-gender lesbian/gay and straight people (e.g., all female targets should indicate their relationships to lesbians and straight women) based on Schubert and Otten (2002). A smaller circle represents the self and a larger circle represents the overall group. From picture one to seven, circles approximated each other, overlapped, and were finally congruent with regard to their centers. Scores for "self – lesbians/gay men" descriptions were recoded so that high values represent a high degree of heteronormative affiliation.

Social Environment

Different measures for determining targets' social environment were applied encompassing circle of friends, current contact to same-gender lesbian/gay and straight people, and contact to boys and girls during childhood.

Circle of friends. In order to represent targets' social environment, we measured gender distribution of friends and sexual orientation of same-gender friends on two separate 7-point scales. Targets completed the statement "My circle of friends consists of..." ranging from 1 (*...men*) to 7 (*...women*) for the gender distribution of friends' item and from 1 (*...lesbians/gay men only*) to 7 (*...straight women/men only*) for the sexual orientation of same-gender friends' item. A small correlation could be observed between the two items (r = -.28, p = .052).

Contact to same-gender lesbian/gay and straight people. In addition to the characteristics regarding the circle of friends, frequency of contact to same-gender lesbian/gay men and straight people (e.g., contact to lesbians and straight women for female targets) should be evaluated on two independent 7-point-scales ranging from 1 (*never*) to 7 (*always*). Both items (contact to lesbian/people and contact to straight people) did not correlate significantly ($r = .05$, $p = .75$).

Contact to girls and boys during childhood. Targets retrospectively evaluated frequency of contact to girls and boys during childhood on two separated scales ranging from 1 (*never*) to 7 (*always*).

Procedure

Production

Data collection for targets was done in two phases. First, targets filled out an online questionnaire that comprised psychological measures. This approach served to ensure similar numbers of lesbian/gay and straight targets. In the second step, speech recordings were gathered in a sound-treated laboratory either by a mid-twenty year old male experimenter for male targets or by a mid-twenty year old female experimenter for female targets. Targets received 18 € for their participation. For more details on the procedure chosen for target data collection, please see *reference deleted for anonymous review*.

Perception

All raters were tested in individual cubicles in a quiet room either in town or on campus. The experimenter welcomed each rater and gave a short description of the experiment including its duration and purpose. In the cubicle, after signing informed consent and putting headphones on, they were asked to follow the instructions on the computer screen. Raters were told that the computer would present photos and/or voice samples of people differing in sexual orientation, one by one, and that their task was to judge targets' sexual orientation as either lesbian/gay or straight. They were instructed to follow their intuition and answer as fast as possible. Then they went through the exercise block. Afterwards, raters read another instruction that again mentioned making ratings intuitively. By pressing the space bar, the experimental blocks started. At the end of the rating task, demographics were collected. Then, raters were thanked and dismissed. The whole experiment lasted about 15-20 minutes.

The experiment was developed and presented using PsyScope X B 53 (Cohen, MacWhinney, Flatt, & Provost, 1993) on MacBooks. The experimental task consisted of 36 experimental blocks, each consisting of 6 trials. In two trials within each block photographs were presented (one of a male face and one of a female face), in two we presented voice recordings (one male and one female), and two were combined trials consisting of photographs and voice recordings (one of a male voice and the corresponding face and one of a female voice and the corresponding face). Each target was presented three times across the experiment, once in each signal condition (voice, face, and voice+face). Stimuli appeared in an individually randomized order for each rater. Each rater received the experimental blocks in the same order while the trials within the block were presented randomly.

The photographs were presented on a white background in the middle of the computer screen. If no photograph was shown, only the voice recording could be heard while the screen remained white. We assessed sexual orientation ratings using a two-alternative forced choice, gay/lesbian versus straight. Raters were asked to press the key “e” if they believed the target was straight, and to press the key “o” for rating the target as gay or lesbian. As a reminder both options were presented in the lower left and right screen corners. Immediately after pressing one of the keys, the next trial started. As a compensation, they received either 4 € or course credit.

Design

Main dependent variables were ratings of targets. Independent variables were signal condition (voice vs. face vs. voice+face), target gender, and target sexual orientation. For targets as the units of analysis, signal condition was a within factor and target gender and sexual orientation were between factors; for raters as the units of analysis, all independent variables were within subject. In all analyses in the present article, significance tests were conducted with $\alpha \leq .05$.

Results

Preliminary Analyses

Across each condition (target gender × target sexual orientation × signal condition), sexual orientation ratings of female and male raters were significantly correlated. Correlations ranged from $r = .70$ (ratings for straight women's voices) to $r = .97$ (ratings for lesbians' voices+faces). Mean correlation was $r_M = .86$. Hence, female and male raters agreed on judgments about sexual orientation in different conditions.

Internal consistencies for all conditions ranged between $\alpha = .77$ (for ratings of straight women's voices) and $\alpha = .98$ (for ratings of lesbians' faces). Interrater reliability across all conditions was $\alpha = .97$. Thus, raters displayed very similar sexual orientation perceptions of a given target. Consequently, we computed perceived sexual orientation scores for each target across raters.

Implicit Signals

A first set of analyses compared accuracies of sexual orientation detection from vocal, facial, and combined signals for female and male targets. Subsequently, a second set of analyses regards specific characteristics inherent in vocal signals and test a first mediation on EPSOM's indirect route (namely, whether the effect of actual on perceived sexual orientation is mediated by acoustic parameters).

Comparison of Signals

A precondition for testing EPSOM's indirect path is that actual sexual orientation can be perceived when presenting different implicit signals with an above-chance accuracy. Moreover, we test whether voices are particularly revealing compared to faces and we ask whether the combination of voices and faces lead to higher accuracies than the single signals and whether there are any differences due to target gender using signal detection analysis. Finally, by using correlation analysis, we tested whether judgments based on one signal can be used to deduce judgments based on another signal (i.e., is a person who is seen as straight, heard as straight?).

Signal Detection Analysis by Signal Condition and Target Gender

In line with previous psychological research on the perception of sexual orientation, we applied signal detection analysis to our data. If a rater categorized a straight target as straight, it was taken as a hit, if a rater categorized a lesbian/gay target as straight, it was coded as a false alarm. We computed the hit- and false-alarm-rate for every signal condition by dividing the sum of hits and false alarms by the total number of straight and lesbian/gay targets in each signal condition. Subsequently, we computed bias-free sensitivity d' representing the difference of standardized hit- and false alarm rates and standardized response bias C representing the criterion to rate targets as either lesbian/gay or straight (Macmillan, 2002). Thus, d' of 0 indicated that raters were only guessing sexual orientation and the more positive values d' takes, the higher the number of straight targets who were correctly classified as straight. A response bias C of 0 indicates no bias, the more positive values C takes, the higher the number of straight judgements (see Online Appendix for analysis and Figure A2 depicting results). A 3 × 2 ANOVA with raters as the units of analysis was computed, with the within-subject factors signal condition and target gender.

Results for d' are depicted in Figure 2 and showed that in every condition female and male targets were identified more correctly than by simple guessing: all $t_{Female}(35) = 4.95$, all $p < .001$, and all $t_{Male}(35) = 3.59$, all $p < .001$. We found an interaction of signal condition and target gender, $F(2, 99) = 24.93$, $p < .001$, $\eta^2_p = .34$, in addition to a main effect of signal condition, $F(2, 99) = 112.83$, $p < .001$, $\eta^2_p = .70$. Simple-effects tests with Bonferroni adjustment showed the following findings. Whereas for male targets discrimination increased from voice over face to their combination, $F(2, 99) = 14.97$, $p < .001$, $\eta^2_p = .23$, for female targets discrimination for face and voice+face was better compared to voice condition, $F(2, 99) = 25.53$, $p < .001$, $\eta^2_p = .34$. Moreover, female targets were identified more correctly than male targets only in face condition, $F(1, 100) = 9.26$, $p = .003$, $\eta^2_p > .09$, but not in voice and combination condition, both $F(1, 100) \leq 1.44$, both $p \geq .22$, both $\eta^2_p \leq .01$. Taken together, raters perceived targets' sexual orientation with above chance accuracy in voice, face and combination condition independent of target gender. Combined signals led to higher accuracies than each single signal for male but not female targets.

Correlations of Judgments regarding Different Signals

Can judgments on one signal be used to deduce judgments based on another signal? Judgments about sexual orientation were related to one another (see Table 1). Thus, the straighter a person was heard, the straighter she or he was seen, and vice versa ($r = .35$, $p = .003$; for correlations on gender-role conformity see Online Appendix). Interestingly, correlations of straight judgments for voice and combination condition ($r = .54$, $p < .001$) were lower than for face and combination condition ($r = .96$, $p < .001$; all $|z| \geq 6.29$, all $p < .001$). The result pattern suggests that overall impressions about perceived straightness were primarily based on faces in contrast to voices. This is supported by a stepwise regression analysis of perceived straightness in combination condition on perceived straightness in voice and in face condition (see Table 3): Face condition was a much stronger predictor than voice condition (face condition: $\beta = .88$, $p < .001$; voice condition: $\beta = .23$, $p < .001$), although voice condition additionally explained variance in combination condition, R^2*-change* = .05, $F(1, 69) = 84.23$, $p < .001$. Hence, when raters were simultaneously given different signals carrying sexual orientation (e.g., voices and faces), the overall raters' impression of targets' sexual orientation is primarily based on faces instead of voices.

However, when examining if this pattern occurred independent of targets' gender, gender-separated regression analyses showed that this was true especially for female targets (see Table 2). Face was a far stronger predictor for combined signals than voice in female targets than in male targets (face-voice difference for female targets: $\Delta\beta = .86$; face-voice difference for male targets: $\Delta\beta = .65$). Hence, when comparing different implicit signals proposed by EPSOM, additional sexual orientation information can be derived for male compared to female targets when their voices were presented synchronously with their faces which is in line with the above findings from signal detection analysis. Moreover, the supremacy of faces compared to voices in signaling sexual orientation is evident for female and male targets. For German targets, we can clearly reject the idea that voices are particularly revealing, as compared to faces.

Acoustic Characteristics

A second set of analyses was used to answer questions on specific implicit signals inherent in voices. When the suitability of EPSOM's prediction that acoustic parameters mediate the effect of actual on perceived sexual orientation should be tested, a precondition is that some acoustic parameters are be connected to perceived sexual orientation in women and men in conditions that include voice. Providing such evidence seems to be particularly necessary given that voices were less reliable signals compared to faces. Hence, we wanted to identify the (most important) acoustic correlates of perceived sexual orientation and whether they are able to mediate the effect of actual on perceived sexual orientation.

Correlations of Acoustic Characteristics and Perceived Straightness

Given the results of the signal detection analysis where voices were found to contain information additionally relevant for correctly judging the sexual orientation of male compared to female targets while presenting faces, one would expect acoustic parameters to signal sexual orientation especially for men. However, acoustic correlates of perceived sexual orientation in voice condition were especially found for women (Table 3). Women were perceived as straighter based on their voices the more gender conforming acoustic characteristics they showed: more expanded vowel spaces, more vowel space shifts to the bottom (mean F1, higher F1 in /aː/, and higher F1 in /ɔ/) and to the front of the mouth (mean F2, higher F2 in /aː/, higher F2 in /ɪ/), the higher their mean fundamental frequency, and the higher their lower fundamental frequency boundary. Even though judgments in combination condition were mainly based on faces, acoustic parameters correlated with perceived sexual orientation in combination condition as well: Women were perceived as straighter in the combination condition, the more expanded their vowel space and the higher their F1 and F2 in /aː/.

Although voice seems to be an additionally relevant signal for males' compared to females' sexual orientation, only two acoustic parameters were linked to perceived straightness in male voices: Male targets were perceived as straighter, the higher their F1 in /ɔ/ – which is an unexpected finding given gender inversion theory (Kite & Deaux, 1987) –

and the lower their lower fundamental frequency boundary. No acoustic parameter was significantly correlated to perceived straightness based on combination condition. Given the number of tests undertaken, significant results for F1 in /ɔ/ and lower fundamental frequency boundary could be chance findings. Hence, it seems that there are acoustic features different from fundamental frequency and vowel space characteristics relevant in men's voice signaling their sexual orientation.

Regression of Perceived Straightness on Acoustic Characteristics

Which acoustic parameters are most important for perceived sexual orientation in voice-including conditions? Because no convincing acoustic correlate was found for men in the voice-only condition, we computed a regression analysis for female targets only. Perceived sexual orientation was regressed on acoustic parameters that were found to correlate with perceived sexual orientation depending on voice condition (vowel space expansion, mean F1 and F2, F1 and F2 in /aː/, F2 in /ɪ/, F1 in /ɔ/, mean fundamental frequency, and lower fundamental frequency boundary), and voice+face condition (vowel space expansion, F1 and F2 in /aː/). Hence, separate regression models for voices and voices+faces were computed using stepwise inclusion of acoustic parameters.

For voice condition, the overall regression model was significant, $F(2, 33) = 15.68$, $p \leq .001$, *adj.* $R^2 = .46$. Female targets were perceived as straighter based on voices only, the higher F1 in /aː/ ($B = .001$, $SE = .000$, $\beta = .60$, $p < .001$) and the higher mean fundamental frequency ($B = .001$, $SE = .001$, $\beta = .30$, $p = .024$). F1 in /aː/ was twice as important as mean fundamental frequency.

For combination condition, the overall regression model was significant as well, $F(1, 34) = 12.12$, $p = .001$, *adj.* $R^2 = .24$. Female targets were perceived as straighter, the higher F1 in /aː/ ($B = .001$, $SE = .000$, $\beta = .51$, $p = .001$). Consequently, regarding specific characteristics of implicit signals proposed by EPSOM, F1 in /a:/ is the most important voice parameter in explaining perceived sexual orientation of female targets independent of signal condition.

Mediation of the Effect of Actual on Perceived Sexual Orientation by Acoustic Characteristics

In order to test indirect effects of actual on perceived sexual orientation via any acoustic parameters gathered, we used the regression-based approach proposed by Hayes (2013). In all mediation analyses in the present paper, we used 10,000 bootstrapping-resamples and 95% bias-corrected confidence intervals (CI) for the estimation of the indirect effects (Preacher & Hayes, 2008). An effect was taken as significant, when the confidence interval did not span zero. We computed separate analyses for each possible acoustic mediator in order to give the analyses adequate power.

For female targets, F1 in /a:/ (b = .01, SE < .01, CI [.00, .02]) and F2 in /a:/ (b = .01, SE < .01, CI [.00, .01]) were found to fully mediate the relationship between actual and perceived sexual orientation in voice condition (for both direct effects: both b = .01, both SE ≤ .01, no CI spanned 0): The straighter female targets described themselves, the more likely they realized gender conforming acoustic parameters (higher F1 and F2 in /a:/), the more likely they were judged as straight based on their voices. When entering F1 and F2 in /a:/ in one mediation model, only F1 in /a:/ fully mediated the relationship between actual and perceived sexual orientation (see Figure 3). Additionally, F1 in /a:/ was shown to partially mediate the relationship in combination condition in the expected way as well (b = .02, SE ≤ .01, CI [.00, .04]; direct effect: b = .05, SE ≤ .02, CI [.01, .08]).

For male targets, there were no significant indirect effects of acoustic parameters mediating the relationship between actual and perceived sexual orientation (all b < .01, SE > .00, change of signs for all CIs). Taken together, these findings support the conclusion drawn from correlation and regression analyses. For female targets, F1 in /a:/ was shown to be the most important acoustic parameter mediating the relationship between actual and perceived sexual orientation. Hence, there is first evidence of an indirect route transmitting actual to perceived sexual orientation as proposed by EPSOM. For male targets, acoustic parameters different from the ones collected in the present study seem to be relevant.

Psychological Characteristics

In order to answer the question which psychological target characteristics are associated with a perception as straight, we used correlation analysis, giving us the opportunity to pre-select the most important psychological characteristics for regression analyses, which is advisable considering the relation of number of targets and number of predictors. Afterwards, we test whether psychological characteristics mediate the relationship between actual and perceived sexual orientation.

Correlations of Psychological Characteristics and Perceived Straightness

Correlation analysis was done for the overall sample as well as separately for the female and male sub-sample, respectively, in every signal condition. Directions of all significant correlations were the same: Perceived straightness was accompanied by gender conforming and heteronormative scores (see Table 4). For face and voice+face condition an almost identical pattern of significant psychological correlates was observed: Psychological characteristics that were significant in face condition were significant in voice+face condition, too.

One psychological characteristic out of each of the three classes of psychological characteristics turned out to be a correlate of perceived straightness: TMF as a measure of actual gender-role conformity, group affiliation to same-gender straight people as a social group affiliation measure, and sexual orientation of same-gender friends as a social environment measure. In general, female and male targets were perceived as straighter the more gender conforming they described themselves, the more group affiliation to same-gender straight people they reported, and the straighter their same-gender friends. All three psychological correlates were associated with perceived straightness in each signal condition and for almost all samples (overall sample, female sub-sample, and male sub-sample). Additionally, group-affiliation to lesbians and gay men was found to be important for all samples but only for face-including conditions.

All other significant psychological correlates were observed mainly for face-including conditions and were due to male targets. Male targets were perceived as straighter in face-including conditions the more gender-conforming (GEPAQ-F, GRB-F, CGRC) and heteronormative their scores were (contact to boys during childhood, gender distribution of friends, contact to gay men and straight men).

Regression of Perceived Straightness on Psychological Characteristics

In order to test which of the psychological characteristics were most relevant in predicting perceived straightness, we used regression analyses. Because of the limited number of targets, not all of the psychological characteristics measured could be taken into account when explaining which of them most affect perceived straightness. In order to preserve appropriate relations of number of targets and number of predictors, we chose the three that showed significant correlations for each signal condition and for almost all samples. Additionally, we included straight attraction index because signal detection analysis suggested its importance. Consequently, we did gender-separate regression analyses for each signal condition by entering straight attraction index, Traditional Masculinity-Femininity Scale

(TMF), group affiliation to same-gender straight people, and sexual orientation of same-gender friends using stepwise inclusion of predictors.

For men, independent of signal condition, the overall regression models were significant, all $F(1, 34) = 6.82$, all $p \leq .013$, all *adj.* $R^2 \geq .14$. Sexual orientation of male targets' male friends was the only significant predictor (all $B \geq .03$, all $SE \leq .01$, all $\beta \geq .41$, all $p \leq .013$). The straighter male targets' male friends, the more likely male targets were perceived as straight in every signal condition (other psychological characteristics: all $\beta \leq .25$, all $p \geq .10$).

For women, independent of signal condition, the overall regression models were also significant, all $F(1, 34) = 7.41$, all $p \leq .010$, all *adj.* $R^2 \geq .16$. The more affiliation female targets reported to straight women, the more likely they were perceived as straight in every signal condition (all $B \geq .02$, all $SE \leq .01$, all $\beta \geq .42$, all $p \leq .010$). In face-including conditions, TMF showed an additional influence on perceived sexual orientation: The more gender conforming female targets described themselves, the more likely they were perceived as straight in face and voice+face condition (all $B \geq .10$, all $SE \leq .04$, all $\beta \geq .38$, all $p \leq .012$).

Hence, the "Psychological Characteristics" component proposed by EPSOM showed different patterns for female and male targets. Whereas a social environment characteristic predicted perceived sexual orientation best for male targets, a social group affiliation and a gender-role conformity measure predicted it best for female targets.

Mediation of the Effect of Actual and Perceived Sexual Orientation by Psychological Characteristics

We tested indirect effects of any acoustic parameter measured for mediating the relationship between actual and perceived sexual orientation. For female targets, actual gender-role conformity on the Traditional Masculinity-Femininity Scale (TMF) displayed a significant indirect effect in face-including conditions: The straighter female targets described themselves, the higher their self-rated gender-role conformity TMF, which in turn led to a straighter perception by others in face (b = .02, SE = .01, CI [.01, .05]) and voice+face condition (b = .02, SE = .01, CI [.00, .05]). However, the direct effects were significant, too (all $b \geq .04$, $SE \leq .02$, no CI spanned 0).

For male targets, the indirect effect of actual gender-role conformity in TMF was significant for voice+face condition only (b = 0.01, SE = 0.01, CI [.00, .03]). Additionally, gender conforming contact to boys during childhood showed a significant indirect effect in face-including conditions: The straighter male targets described themselves, the higher their reported contact to boys during childhood which in turn led to a straighter perception based on faces (b = 0.11, SE = 0.01, CI [.00, .03]) and voices+faces (b = 0.14, SE = 0.01, CI [.00, .03]). However, the direct effects were significant, too (all $b \geq .02$, all $SE \leq .01$, no CI spanned 0).

Hence, there is evidence for another indirect route mediating actual to perceived sexual orientation proposed by EPSOM. For female and male targets, actual and perceived sexual orientation is partially mediated by actual gender-role conformity.

Mediation of Effect of Actual on Perceived Sexual Orientation via Psychological and Acoustic Characteristics

We examined the indirect route proposed by EPSOM more comprehensively in a serial mediation analysis. Although EPSOM suggests a clear direction of the indirect route (actual sexual orientation influences psychological characteristics of the targets that in turn affect certain acoustic parameters that lead to a perception as lesbian/gay or straight), mediation analysis does not specify a certain direction. We applied mediation analysis to female targets only, because no acoustic correlates of perceived sexual orientation were found for men. Moreover, we used the Traditional Masculinity-Femininity Scale (TMF) representing psychological target characteristics, because it was shown to be the only psychological characteristic mediating the effect of actual on perceived sexual orientation. Hence, also the opposite direction would be possible.

TMF showed combined indirect effects with a range of acoustic parameters fully mediating the relationship of actual and perceived sexual orientation in voice condition. The straighter female targets described themselves, the more gender conforming they rated themselves on TMF, the more gender conforming acoustic parameters they produced (i.e., higher values for F1 in /aː/, F2 in /ɪ/, mean F1, mean fundamental frequency), which led to a straighter perception by others (all $b \geq .003$, $SE \leq .002$, no CI spanned 0). None of the direct effects were significant (all $b \leq .011$, $SE \geq .006$, all CIs spanned 0).

Discussion

EPSOM suggests three components mediating the relationship between actual and perceived sexual orientation. Focusing on the "Implicit Signals" component, we provided evidence that targets were categorized more accurately and were taken as straight more often based on voices, faces, and combined signals than could be expected by simple guessing. Faces were a better signal to female and male sexual orientation than voices. However, voices provided additional information on male but not female sexual orientation information when they were presented along with faces compared to faces alone. Moreover, we found that sexual orientation judgments between different signals were related: When a target person was seen as straight, she or he was heard as straight. Supporting the findings from signal detection analysis, judgments based on faces predicted judgments based on both signals better for female compared to male targets. Hence, expression and perception of female sexual orientation seems to be primarily face-based whereas male sexual orientation includes vocal information to some degree. Consequently, it was expected that when focusing on specific characteristics inherent in vocal signals more acoustic correlates of perceived sexual orientation of men than women could be shown. However, raters used acoustic parameters to determine female sexual orientation only: F1 in /aː/ was found to be the most important acoustic correlate of perceived sexual orientation and the most important mediator of the relationship between actual and perceived sexual orientation.

Regarding EPSOMs „Psychological Characteristics” component, characteristics from all three classes were shown to be psychological correlates of perceived sexual orientation independent of signal condition and target gender in line with expectations (i.e., the more gender conforming and heteronormative targets described themselves, the more likely they were perceived as straight). However, the importance of the psychological correlates differed based on gender. Whereas actual sexual orientation was most important for men, affiliation to straight women best predicted perceived sexual orientation for women in every signal condition. Additionally, actual gender-role conformity in face-including conditions was important for perceiving female targets’ sexual orientation. Hence, there is suggestive evidence for another indirect route mediating the effect of actual on perceived sexual orientation proposed by EPSOM, at least for female targets. Actual gender-role conformity partially mediated the relationship of actual on perceived sexual orientation for female and male targets. In a more comprehensive analysis, the straighter female targets described themselves, the more gender conforming they rated themselves, which, in turn, led to more gender conforming acoustic parameters and effected a straighter perception by others. Hence, as proposed by EPSOM, psychological characteristics and implicit signals mediated the relationship of actual and perceived sexual orientation in combination. What is yet missing is data on the "Impressions Related to Sexual Orientation" as the third component proposed by EPSOM.

Experiment 2

The main aim of Experiment 2 was to provide further evidence on EPSOM's indirect route and its mediators with a special focus on impressions related to sexual orientation. Hence, we gathered data on perceived gender-role conformity based on voices and faces. First, we tested how actual and perceived gender-role conformity were connected to each other when presenting the two signals and which were the acoustic correlates of perceived gender-role conformity. Thereby we further tested EPSOM's "Implicit Signals" component. Second, we asked which psychological characteristics lead to a perception as gender conforming. Third, we tested if perceived gender-role conformity was an additional mediator (in combination with actual gender-role conformity and acoustic parameters) of the relationship between actual and perceived sexual orientation, hence fully testing the indirect route proposed by EPSOM.

Method

The same voice and face stimuli from the 18 targets per group (lesbians, gay men, straight men and women) as in Experiment 1 were used, except for one lesbian target who was replaced by another lesbian target for technical reasons. Hence, data for targets' psychological characteristics (sexual orientation, actual gender-role conformity, social group affiliation, and social environment), acoustic parameters, and perceived straightness were almost the same as in Experiment 1. The straight attraction index was used as a measure of actual sexual orientation and the Traditional Masculinity-Femininity Scale (TMF) was used as a measure of actual gender-role conformity (both scales ranging from 1 to 7).

Overall, 38 participants completed the experiment. Participants were asked to rate all target voices and faces on a 7-point a masculinity/femininity scale ranging from 1 (*totally masculine*) to 7 (*totally feminine*); ratings were recoded in order to have higher scores indicate higher perceived gender-role conformity. They were randomly assigned to two different orders of blockwise presentation (1^{st}: male faces, male voices, female faces, female voices; 2^{nd}: male voices, female voices, male faces, female faces). Stimuli were randomly presented within each block. In contrast to Experiment 1, there was no combination condition.

Results

Preliminary Analysis

Internal consistency for all conditions (target gender × target sexual orientation × signal condition) ranged between $\alpha = .94$ (for ratings of straight men's faces) and $\alpha = .98$ (for ratings of lesbians' faces). Interrater reliability across all conditions was $\alpha = .97$. Thus, raters displayed very similar gender-role conformity perceptions of a given target. Consequently, we computed perceived gender-role conformity scores for each target across raters.

Implicit Signals

Focusing on EPSOM's "Implicit Signals"-component, a first set of analyses refers to whether raters derive female and male targets' gender-role conformity better from vocal vs. facial signals. Subsequently, a second set of analyses regards specific characteristics inherent in vocal signals and test a first mediation on EPSOM's indirect route including "Impressions Related to Sexual Orientation" (namely, whether the effect of actual on perceived gender-role conformity is transmitted by acoustic parameters).

Comparison of Signals

Are actual and perceived gender-role conformity associated with each other in the different signals? And can gender-role conformity be deduced better when based on voices or on faces? As correlation analyses showed, the more gender conforming targets described themselves, the more gender conforming they were perceived by others based on voices (r = .32, p = .006) and faces (r = .49, p < .001; see Table 1). Hence, gender-role conformity is expressed by different signals and there was no significant difference for voices and faces providing information on gender-role conformity in the overall sample (z = -1.40, p = .081).

However, correlations of actual and perceived gender-conformity for voice and face condition were different for male targets (r_{voice} = .14, r_{face} = .43, z = 1.83, p = .034) but similar for female targets (r_{voice} = .53, r_{face} = .56, z = -.19, p = .423). Moreover, correlations of actual and perceived gender-role conformity for female and male targets were only similar in face condition (r_{female} = .56, r_{male} = .43, z = .70, p = .241) but different in voice condition (r_{female} = .53, r_{male} = .14, z = 1.83, p = .034). Thus, voice seems to be a less valid signal than face for male targets' but not for female targets' gender conformity.

Acoustic Characteristics

Which acoustic parameters are connected to perceived gender-role conformity in women and men in voice condition? Do acoustic parameters mediate the effect of actual on perceived gender-role conformity?

Correlations of Acoustic Characteristics and Perceived Gender-Role Conformity

Which acoustic characteristics are associated with perceived gender-role conformity? Given the previous results that voice was a weak signal for male targets' gender-role conformity, we expected null findings for male targets. However, male targets were perceived as more gender conforming the lower their lower fundamental frequency boundary, mean fundamental frequency, and F2 in /aː/. In contrast to perceived sexual orientation and contrary to expectations, male targets' acoustic correlates of perceived gender-role conformity were clearer. Female targets were perceived as more gender conforming based on their voices the more expanded their vowel spaces, the more their vowel spaces were directed to the bottom (mean F1, higher F1 in /aː/) and to the front of the mouth (mean F2, higher F2 in /aː/, higher F2 in /ɪ/), and the higher their mean fundamental frequency (see Table 3). These findings parallel those of perceived sexual orientation of female targets (except for F1 in /ɔ/ and lower fundamental frequency boundary that are not shown to be acoustic correlates of perceived gender-role conformity). Taken together, while the acoustic pattern of perceived gender-role conformity almost mirrored that of perceived sexual orientation for female targets, the pattern was more convincing for male targets. Hence, further analyses referred to male in addition to female targets' acoustic parameters.

Regression of Perceived Gender-Role Conformity on Acoustic Characteristics

To test which acoustic parameters are most important for perceived gender-role conformity in voice condition, we regressed perceived gender-role conformity based on voices on acoustic parameters that were found to correlate with perceived gender-role conformity for female (vowel space expansion, mean F1 and F2, F1 and F2 in /aː/, F2 in /ɪ/, mean fundamental frequency) and male targets (vowel space expansion, F1 and F2 in /aː/). Hence, separate regression models using the stepwise inclusion of acoustic parameters were computed for female and male targets.

For female targets in voice condition, the overall regression model was significant, $F(3, 31) = 23.78$, $p \leq .001$, *adj.* $R^2 = .67$. Female targets were perceived as more gender conforming based on voices, the higher mean F2 (B = .005, SE = .001, β = .49, p < .001), mean fundamental frequency (B = .025, SE = .005, β = .45, p < .001), and F1 in /a:/ (B = .004, SE = .001, β = .33, p = .005).

For male targets, the overall regression model was significant as well, $F(2, 33)$ = 15.42, p < .001, *adj.* R^2 = .45. Male targets were perceived as more gender conforming based on voices, the lower their lower fundamental frequency boundary (B = -.053, SE = .011, β = -.59, p < .001) and F2 in /a:/ (B = -.002, SE = .001, β = -.29, p = .030). Lower fundamental frequency boundary was twice as important as F2 in /a:/.

Mediation of the Effect of Actual and Perceived Gender-Role Conformity by Acoustic Characteristics

For female targets, F1 in /a:/ (b = .17, SE = .08, CIs [.00, .02]), mean F1 (b = .15, SE = .10, CIs [.03, .42]), and mean fundamental frequency (b = .13, SE = .09, CIs [.02, .40]) were found to partially mediate the relationship between actual and perceived gender-role conformity in voice condition (for all direct effects: all b > .35, all SE < .16, no CI spanned 0). The more gender conforming female targets described themselves, the more gender conforming acoustic characteristics they produced (higher F1 in /a:/, mean F1, and mean fundamental frequency), and the more likely they were perceived as gender conforming. When entering F1 in /a:/, mean F1, and mean fundamental frequency in one mediation model, only mean fundamental frequency showed a significant indirect effect and fully mediated the relationship between actual and perceived gender-role conformity (b = .17, SE = .08, CIs [.05, .39]; indirect effects for F1 in /a:/ and mean F1: b < .12, SE > .11, all CIs spanned 0; direct effect: b = .14, SE = .16, CIs [-.19, .46]).

For male targets, mean fundamental frequency showed the only significant indirect effect partially mediating the relationship between actual and perceived gender-role conformity. The more gender conforming male targets described themselves, the lower their mean fundamental frequency, and the more likely they were perceived as gender conforming (b = -.38, SE = .13, CI [-.68, -.15]; direct effect: b = .52, SE = .15, CI [.22, .82]).

Psychological Characteristics

Focusing on EPSOM's "Psychological Characteristics" component, we ask which psychological target characteristics lead to a perception as gender conforming. In line with analyses for sexual orientation, we used correlation analyses first in order to pre-select the most important psychological characteristics for regression analyses.

Correlations of Psychological Characteristics and Perceived Gender-Role Conformity

As for perceived sexual orientation, correlation analyses were done for the overall sample and separately for the female and male sub-sample in voice and face condition. Significance levels were left unadjusted for exploratory reasons again. Directions of all significant correlations pointed in the same direction: Perceived gender-role conformity was accompanied by gender-conform and heteronormative scores on psychological characteristics (see Table 3).

Those psychological correlates that were associated with perceived gender-role conformity in each signal condition and almost for all samples (overall sample, female sub-sample, and male sub-sample) were considered as most important again. According to that criterion, straight attraction index, Traditional Masculinity-Femininity Scale (TMF), group affiliation to same-gender straight people, and sexual orientation of same-gender friends turned out to be important correlates of perceived gender-role conformity. Additionally, Feminine Gender-Role Conforming Behavior (GRB-F) and Childhood Gender-Role Conformity (CGRC) showed significant correlations with perceived gender-role conformity.

Although one characteristic from each of the four psychological classes was significant, gender-related measures played the most important role in accompanying perceived gender-role conformity.

Regression of Perceived Gender-Role Conformity on Psychological Characteristics

In order to test which psychological target characteristics were most important in predicting perceived gender-role conformity, we conducted gender-separate regression analyses for voice and face condition by entering the psychological correlates found in the previous correlation analyses (straight attraction index, TMF, GRB-F, CGRC, group affiliation to same-gender straight people, sexual orientation of same-gender friends) using stepwise inclusion of predictors.

For men, independent of signal condition, the overall regression models were significant, both $F(1, 34) \geq 8.83$, both $p \leq .005$, both *adj.* $R^2 \geq .18$. The straighter male targets described themselves, the more gender conforming they were perceived by raters in voice and face condition (both $B \geq .18$, both $SE \leq .06$, both $\beta \geq .45$, both $p \leq .005$). Hence, only actual sexual orientation influenced how men's sexual orientation was perceived.

For female targets, the overall regression models were significant in voice and face condition, both $F(2, 32) \geq 11.67$, both $p < .001$, both *adj.* $R^2 \geq .39$. Compared to male targets, the prediction of perceived gender-role conformity resulted in a more complex picture. In face condition, perceived gender-role conformity was explained by actual gender-role conformity on TMF ($B = .47$, $SE = .17$, $\beta = .41$, $p = .009$) and by group affiliation to straight women ($B = .26$, $SE = .11$, $\beta = .36$, $p = .021$): The more gender conforming female targets described themselves and the more affiliated to straight women they felt, the more likely they were to be rated as gender conforming by others in face condition. In voice condition, actual gender-role conformity on CGRC (B = .55, SE = .13, β = .56, p < .001) and sexual orientation of female friends (B = .23, SE = .09, β = .34, p = .014) were most important. The more gender conforming female targets described themselves and the straighter their female friends, the more likely their voices were perceived as gender conforming.

Taken together, while perceived gender-role conformity was predicted best by actual sexual orientation for male targets, for female targets actual gender-role conformity in combination with an actual sexual orientation related characteristics (e.g., sexual orientation of same-gender friends and group affiliation to same-gender straight people) were most important in explaining perceived gender-role conformity.

Actual and Perceived Sexual Orientation and Gender-Role Conformity

We wanted to expand evidence for EPSOM's indirect route, using mediation analysis.

Mediation of the Effect of Actual on Perceived Sexual Orientation by Perceived Gender-Role Conformity

In order to further test EPSOM's indirect route, we examined whether there is an indirect effect of actual sexual orientation via perceived gender-role conformity on perceived sexual orientation. Hence, we computed mediation analyses in order to determine whether perceived gender-role conformity mediated the relationship between actual and perceived sexual orientation for voice and face condition, respectively, using the same settings as described above. Perceived gender-role conformity fully mediated the relationship between actual and perceived sexual orientation in voice, b = .01, SE = .00, CI [.00, .03], and face condition, b = .04, SE = .00, CI [.02, .06], because the direct effects were not significant (both $b \leq .01$, both $SE \geq .01$, both CIs spanned 0). The same pattern occurred when conducting gender-separate analyses for voice and face condition, respectively (all indirect effects: b >

.01, $SE \leq .01$, no CI spanned 0; all direct effects: $b < .01$, $SE > .00$, no CI spanned 0). Hence, it can be concluded that the straighter targets described themselves, the more gender conforming they were perceived by raters which in turn led to a straighter perception by raters. This indirect effect was obtained independent of targets' gender and signal condition.

Mediation of the Effect of Actual on Perceived Sexual Orientation by Actual and Perceived Gender-Role Conformity

In order to additionally account for psychological characteristics as potential mediators as proposed in EPSOM, we used actual gender-role conformity on Traditional Masculinity-Femininity Scale (TMF) because it was shown to be one of the possible mediators of actual and perceived sexual orientation. Hence, we tested combined indirect effects of actual and perceived gender-role conformity mediating the relationship of actual and perceived sexual orientation for the overall sample as well as the female and male subsample, respectively, and separately for voice and face condition in serial mediation analysis.

Full mediations of combined indirect effects of actual on perceived gender-role conformity were found for face condition independent of the examined sample (all $b > .004$, all $SE < .010$, no CI spanned 0; all direct effects: all $b < .012$, all $SE > .01$, no CI spanned 0). The straighter female and male targets described themselves, the more likely they rated themselves as gender conforming, the more likely they were rated as gender conforming by others in face condition, and the more likely they were to be rated as straight by others based on their faces. Combined indirect effects of actual and perceived gender-role conformity based on voices were found for female targets only ($b = .007$, $SE = .003$, CI [.00, .03]; direct effect: $b = .007$, $SE = .004$, CI [-.002, .016]). As described for face condition, all effects pointed in the expected directions.

Mediation of the Effect of Actual on Perceived Sexual Orientation by Actual Gender-Role Conformity, Acoustic Characteristics, and Perceived Gender-Role Conformity

Finally, we tested EPSOM's overall indirect route by taking all mediating components into account. Was the relationship between actual and perceived sexual orientation mediated by psychological characteristics indicated by actual gender-role conformity on TMF, implicit signals that were represented by diverse acoustic parameters, and impressions related to sexual orientation which were denoted by perceived gender-role conformity? We computed mediation models for perceptions based on voice condition only, because acoustic parameters represented the "Implicit Signals" component. Furthermore, mediation analysis was restricted to female targets, because mediations of actual on perceived sexual orientation via a) acoustic parameters (see section "Mediation of Actual and Perceived Sexual Orientation by Psychological Characteristics") and b) actual and perceived gender-role conformity (see section "Mediation of Actual on Perceived Sexual Orientation by Actual and Perceived Gender-Role Conformity") were found for female targets only.

Full mediations were found including mean fundamental frequency, mean F1, F1 in /a:/, F2 in /ɪ/, and vowel space expansion (all $b \geq .0012$, all $SE \leq .0017$, no CI spanned 0; all direct effects: all $b \leq .0069$, all $SE \geq .0042$, all CIs spanned 0). The straighter female targets described themselves, the more likely they described themselves as gender conforming, the more likely they produced gender conforming acoustic parameters (higher mean fundamental frequency, mean F1, F1 in /a:/, F2 in /ɪ/, and greater vowel space expansion), the more likely they were perceived as gender conforming, and the more likely they were to be perceived as straight.

Discussion

In Experiment 2, we focused on the interrelations of EPSOM's "Impressions Related to Sexual Orientation" component with other mediating components and its mediatory function on the indirect route transmitting actual to perceived sexual orientation. Regarding the "Implicit Signals" component, gender-role conformity is expressed and perceived by different signals, but voice seems to be a weak signal for male targets. Moreover, there were some mostly important acoustic correlates of perceived gender-role conformity for female (mean F2, mean fundamental frequency, and F1 in /a:/) and male targets (fundamental frequency boundary, F2 in /a:/); mean fundamental frequency mediated the relationship between actual and perceived sexual orientation completely for female targets and partially for male targets. Hence, male targets' voices do carry information about their gender-role conformity. Regarding the "Psychological Characteristics"-component, actual sexual orientation was the most important psychological correlate of perceived gender-role conformity for male targets. For female targets, actual gender-role conformity combined with a characteristic linked to actual sexual orientation (sexual orientation of same-gender friends and group affiliation to same-gender straight people) explained perceived gender-role conformity.

There were several pieces of evidence on the mediatory function of the "Impressions Related to Sexual Orientation" component. Perceived gender-role conformity mediated the relationship of actual and perceived sexual orientation when included as a single mediator (for female and male targets based on vocal and facial signals), when included in combination with actual gender-role conformity representing the "Psychological Characteristics" component (except for male voices), and when testing EPSOM's overall indirect route by adding acoustic parameters indicating the "Implicit Signals" component. The straighter female targets described themselves, the more gender conforming they described themselves, the more gender conforming acoustic parameters they used (higher mean fundamental frequency, mean F1, F1 in /aː/, F2 in /ɪ/, and greater vowel space expansion), the more gender conforming they were perceived based on their voices, and the more likely they were to be judged as straight. Hence, several acoustic parameters were able to mediate the sexual orientation signal.

General Discussion

Why and how are people able to judge others' sexual orientation with above-chance accuracy? We introduced the expression and perception of sexual orientation model (EPSOM) that details under which conditions people are able to express their sexual orientation and perceive it from others thereby integrating a majority of findings from corresponding research. Proposing an indirect route, EPSOM refers to three components mediating the relationship between actual and perceived sexual orientation: Actual sexual orientation is related to several *psychological characteristics* (e.g., actual gender-role conformity) which are in turn linked to *implicit signals* (e.g., voices and faces) that are decoded as *impressions related to sexual orientation* (e.g., perceived gender-role conformity) which are used to perceive sexual orientation. We answered open questions on the mediating components and tested their mediatory function in isolation and in combination in two experiments using female and male targets. Raters judged targets' sexual orientation based on voices, faces, and their combination (Experiment 1) and targets' gender-role conformity based on voices and faces (Experiment 2).

Focusing on the "Implicit Signals" component proposed by EPSOM, actual sexual orientation was judged correctly with above chance accuracy for voice, face, and voice+face signals. Hence, our findings supported previous research that faces carry sexual orientation information also in non-U.S. countries (see Valentova & Havlíček, 2013). By using facial photographs taken under standardized conditions in our lab, we provided further evidence that picture inherent quality differences do not account for above chance accuracy in sexual orientation detection (Rule & Ambady, 2008; Rule et al., 2008; but see, Cox et al., 2015) and avoided any ethical concerns because targets consented to participation. Supporting previous evidence, there was a gender difference for facial and combined signals: Faces were better in signaling female compared to male sexual orientation (Tabak & Zayas, 2012; Lyons et al., 2014) and combined stimuli led to an improvement of accuracy compared to single signals for male but not female targets (Rieger et al., 2010). When faces are such a good signal for female sexual orientation that voices contribute no additional information, then acoustic parameters should be unrelated to perceived sexual orientation of women. Hence, future research could shed some light on whether voices and faces indeed contain conflicting information on sexual orientation.

In contrast to previous findings (Valentova & Havlíček, 2013), faces carried more sexual orientation information than voices and judgments about sexual orientation based on one signal could be used to deduce judgments based on another signal, suggesting that sexual orientation information is spread across signals instead of a trade-off. This inconsistency in findings could be explained by a possibly different relative informational richness of vocal and facial stimuli in the studies. Whereas our vocal stimuli were all standardized by recording the same single sentence for every target leaving little space for inter-individual differences (vs. 20s excerpt of a read text used by Valentova & Havlíček, 2013), we used faces with high ecological validity (vs. faces with removed hair style, facial jewelry, and earrings by Valentova & Havlíček, 2013). However, it is unclear how the different signals have to be composed and standardized so that they contain a comparable amount of inter-individual information. Hence, we recommend future research to further investigate implications of relative richness of information on sexual orientation detection in more detail.

Keeping the focus on EPSOM's "Implicit Signals" component, but turning from signal comparisons to signal inherent characteristics, we found different acoustic correlates of perceived sexual orientation and gender-role conformity for female and male targets. Whereas F1 in /a:/ was found to be the most important acoustic correlate of perceived sexual orientation for female targets, no convincing acoustic correlate was shown for male targets. In contrast, more acoustic correlates were shown regarding perceived gender-role conformity for female (mean F2, mean fundamental frequency, and F1 in /a:/) and male targets (fundamental frequency boundary, F2 in /a:/). Although we used sentences instead of series of single words, we conclude, in line with Munson (2007), that aurally perceived sexual orientation and gender-role conformity are two different constructs, because they are cued by different acoustic parameters. Going a step further, our findings suggest that perceived gender-role conformity is better acoustically represented than perceived sexual orientation, which supports the structure of EPSOM by linking implicit signals and perceived sexual orientation using impressions related to sexual orientation.

Turning to EPSOM's "Psychological Characteristics" component, the importance of the psychological correlates differed based on gender. Whereas actual sexual orientation was most important predicting perceived sexual orientation and gender-role conformity for male targets independent of signal condition, the picture was more complex for female targets. Although it depended on perception and signal which specific psychological correlates explained most variance, actual gender-role conformity was generally more important for perceived gender-role conformity than perceived sexual orientation. With few exceptions (Sylva et al., 2010; Mann, 2012), there is no research on which psychological target characteristics lead to judgments as straighter and gender conforming, although there is much of evidence on psychological correlates of actual sexual orientation. Hence, future research considering other possible characteristics would expand knowledge on this question.

Regarding EPSOM's "Impressions Related to Sexual Orientation" component, we showed that information about gender-role conformity was successfully transmitted by female and male targets' faces and female but not male targets' voices. Corresponding to findings on perceived sexual orientation, we provided supporting evidence (Rieger et al., 2010) that targets who were judged by the means of one signal were rated as gender conforming based on another signal.

Beside the single EPSOM components and their interrelations, the core of EPSOM is how actual sexual orientation is transmitted to perceived sexual orientation. We provided several pieces of evidence that psychological characteristics, implicit signals, and impressions related to sexual orientation explained the association of actual and perceived sexual orientation in isolation and combination. This was particularly true for female targets. Most importantly, findings supported the overall indirect route proposed by EPSOM for female targets: Higher self-rated straightness led to higher gender-role conformity which in turn was mirrored by different more gender conforming acoustic parameters (higher mean fundamental frequency, mean F1, F1 in /aː/, F2 in /ɪ/, and greater vowel space expansion) which raters used to draw conclusions about targets' gender-role conformity that was in turn used as an indicator of targets' sexual orientation. The null finding that the "Implicit Signals" component did not mediate actual and perceived sexual orientation for male targets could be due to limiting acoustic parameters to intonational and vowel space characteristics. Although men who described themselves as gender-conforming showed some gender-conforming speech, the stereotype-truth of sexual orientation for vocal signals was found to be low in line with previous research (Kachel et al., 2017a) because actual sexual orientation explained only 12% of perceived sexual orientation in voices. Maybe other parameters commonly discussed in research on acoustic correlates of men's actual and perceived sexual orientation such as features of the /s/-sound, durational measures, or voice quality characteristics (see review by Munson & Babel, 2007) were better suited for signaling men's sexual orientation. Because we tested EPSOM using speech-based evidence only, future research should test whether specific signals inherent in faces or other implicit signals are able to transmit female and male targets' sexual orientation.

The present study contains some limitations. When rating sexual orientation, signals of female and male targets were presented mixed up instead of blockwise. Hence, the concept of sexual orientation which depends on gender was not constantly present in raters' minds but had to be re-actualized over and over again. This possibly accounts for low accuracies and response biases in voice signals. Moreover, our study included targets who were either lesbian/gay or straight. Although faces of bisexual people were judged similar to those of lesbian/gay people (Ding & Rule, 2012), future studies should test whether EPSOM is able to transmit bisexuality. Moreover, given the target sample size, we were able to detect only moderate to large effects. Using a larger sample would allow future studies to detect even small effects.

Taken together, integrating several previous findings, EPSOM may provide a valuable framework for understanding and explaining, how people express their sexual orientation and how it is perceived by others using different implicit signals such as voices on a telephone or facial photographs in social network media.

Acknowledgments

Substantial contributions to the conception or design of the work: SK, MCS, APS; substantial contribution to the acquisition and analysis of the data: SK, MCS, APS; interpretation of data for the work: SK, MCS, APS; Drafting the work or revising it critically for important intellectual content: SK, MCS, APS; Final approval of the version to be published: SK, MCS, APS.

The current research was funded by a grant from the German Research Foundation to the last two authors (DFG STE 938/10-2; FOR 1097). We thank Jörg Dreyer, Susanne Fuchs, and Stefanie Jannedy for providing the phonetic laboratory at the Zentrum für Allgemeine Sprachwissenschaft in Berlin, we thank Monique Beck, Karoline Nestler, Maren Schlehe, Alexander Makosch, Felix Sokoll, Felicia Schuld, Sonja Gaukel, Leonie Backofen, Jule Herrmann, Felix Ludwig, Markus Mahlmann, and Christin Schmidt for help with data collection and data pre-processing. The authors declare no conflicts of financial, personal, or other interests.

References

Ambady, N., Hallahan, M., & Conner, B. (1999). Accuracy of judgments of sexual orientation from thin slices of behavior. *Journal of Personality and Social Psychology,* 77, 538–547.

Athenstaedt, U. (2003). On the content and structure of the gender role self-concept: Including gender-stereotypical behaviors in addition to traits. *Psychology of Women Quarterly,* 27, 309–318.

Baeck, H., Corthals, P., & van Borsel, J. (2011). Pitch characteristics of homosexual males. *Journal of Voice,* 25, 211–214. doi: 10.1016/j.jvoice.2010.10.019.

Biemans, M. A. J. (2000). *Gender variation in voice quality* (Doctoral dissertation, Catholic University Nijmegen, Nijmegen. Netherlands). Retrieved from http://www.lotpublications.nl/Documents/38_fulltext.pdf

Boersma, P., & Weenink, D. (2014). PRAAT: Doing phonetics by computer (version 5.3.71) [computer code]. Retrieved from http://www.praat.org/ (Last viewed September 18, 2014).

Bosson, J. K., Prewitt-Freilino, J. L., & Taylor, J. N. (2005). Role rigidity: A problem of identity misclassification? *Journal of Personality and Social Psychology*, 89, 552–565.

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication,* 20, 255–272.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, 77, 305–327.

Camp, M. (2011). *Japanese lesbian speech: Sexuality, gender identity, and language* (Doctoral dissertation, University of Arizona). Retrieved from http://hdl.handle.net/10150/195371

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Lawrence Erlbaum Associates, New Jersey), 1–567.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). Psyscope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments & Computers,* 25, 257–271.

Constantinople, A. (1973). Masculinity-femininity: An exception to a famous dictum? *Psychological Bulletin,* 80, 389–407.

Cox, W. T. L., Devine, P. G., Bischmann, A. A., & Hyde, J. S. (2016). Inferences about sexual orientation: The roles of stereotypes, faces, and the gaydar myth. *The Journal of Sex Research,* 53, 157–171. doi: 10.1080/00224499.2015.1015714.

Ding, J. Y. C., & Rule, N. O. (2012). Gay, straight, or somewhere in between: Accuracy and bias in the perception of bisexual faces. *Journal of Nonverbal Behavior,* 36, 165–176.

Dunkle, J. H., & Francis, P. L. (1990). The role of facial masculinity/femininity in the attribution of homosexuality. *Sex Roles,* 23, 157–167.

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, 118, 247–279.

Freeman, J. B., Johnson, K. L., Ambady, N., & Rule, N. O. (2010). Sexual orientation perception involves gendered facial cues. *Personality and Social Psychology Bulletin,* 36, 1318–1331.

Gaudio, R. P. (1994). Sounding gay: Pitch properties in the speech of gay and straight men. *American Speech,* 69, 30–57.

Guzik, K. M. (2004). Acoustic analysis of phonetic parameters of less masculine sounding German speech. *Arbeitsberichte des Instituts für Phonetik und Digitale Sprachverarbeitung Universität Kiel (AIPUK),* 36, 15–29.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis. A regression-based approach*. (Guilford Press: New York)

Hughes, S. M., & Bremme, R. (2011).The effects of facial symmetry and sexually-dimorphic facial proportions on assessments of sexual orientation. *Journal of Social, Evolutionary, and Cultural Psychology,* 5, 214–230.

Johnson, K. L., & Ghavami, N. (2011). At the crossroads of conspicuous and concealable: What race categories communicate about sexual orientation. *PLoS ONE,* 6, S. e18025. doi: 10.1371/journal.pone.0018025

Johnson, K. L., Gill, S., Reichman, V., & Tassinary, L. G. (2007). Swagger, sway, and sexuality: Judging sexual orientation from body motion and morphology. *Journal of Personality and Social Psychology,* 93, 321–334. doi: 10.1037/0022-3514.93.3.321.supp.

Kachel, S., Simpson, A. P., & Steffens, M. C. (2017a). Do I sound straight? Acoustic correlates of actual and perceived sexual orientation and masculinity/femininity in men's speech. Submitted to *Journal of Speech, Language, and Hearing Research.*

Kachel, S., Simpson, A. P., & Steffens, M. C. (2017b). Acoustic correlates of sexual orientation and gender-role self-concept in women's speech. In press at the *Journal of the Acoustical Society of America.*

Kachel, S., Steffens, M. C., & Niedlich, C. (2016). Traditional masculinity and femininity: Validation of a new scale assessing gender roles. *Frontiers in Psychology,* 7:956. doi: 10.3389/fpsyg.2016.00956.

Kite, M. E., & Deaux, K. (1987). Gender belief systems: Homosexuality and the implicit inversion. *Psychology of Women Quarterly,* 11, 83–96. doi: 10.1111/j.1471-6402.1987.tb00776.x.

Kite, M. E., & Whitley, B. E., Jr. (1996). Sex differences in attitudes toward homosexual persons, behaviors, and civil rights: A meta-analysis. *Personality and Social Psychology Bulletin, 22*, 336-353.

Krakauer, I. D., & Rose, S. M. (2002). The impact of group membership on lesbians' physical appearance. *Journal of Lesbian Studies*, 6, 31–43.

Lerman, J., & Damsté, P. (1969). Voice pitch of homosexuals. *Folia Phoniatrica,* 21, 340–346.

Lester, D., McLaughlin, S., Cohen, R., & Dunn, L. (1977). Sex-deviant handwriting, femininity, and homosexuality. *Perceptual and Motor Skills,* 45, 1156. doi: 10.2466/pms.1977.45.3f.1156.

Lick, D. J., & Johnson, K. L. (2013). Deliberate changes to gendered body motion influence basic social perceptions. *Social Cognition,* 31, 656–671.

Lick, D. J., & Johnson, K. L. (2014a). "You can't tell just by looking!": Beliefs in the diagnosticity of visual cues explain response biases in social categorization. *Personality and Social Psychology Bulletin,* 40, 1494–1506. doi: 10.1177/0146167214549323.

Lick, D. J., & Johnson, K. L. (2014b). Perceptual underpinnings of antigay prejudice: Negative evaluations of sexual minority women arise on the basis of gendered facial features. *Personality and Social Psychology Bulletin,* 40, 1178–1192.

Linville, S. E. (1998). Acoustic correlates of perceived versus actual sexual orientation in men's speech. *Folia Phoniatrica et Logopaedica,* 50, 35–48.

Lippa, R. A. (2005). Sexual orientation and personality. *Annual Review of Sex Research,* 16, 119–153.

Loulan, J. (1990). *The lesbian erotic dance: Butch, femme, androgyny, and other rhythms* (Spinsters Ink., Minneapolis), 1-287.

Lyons, M., Lynch, A., Brewer, G., & Bruno, D. (2014). Detection of sexual orientation ("gaydar") by homosexual and heterosexual women. *Archives of Sexual Behavior,* 43, 345–352. doi: 10.1007/s10508-013-0144-7.

Macmillan, N. A. (2002). Signal detection theory. In J. Wixted, & H. Pashler (Eds.), *Stevens' Handbook of Experimental Psychology* (3rd ed., Vol. 4, pp. 43-90). New York: Wiley.

Mann, S. L. (2012). Speaker attitude as a predictive factor in listener perception of gay men's speech. *Journal of Language and Sexuality,* 1, 206–230. doi: 10.1075/jls.1.2.04man.

Munson, B. (2007). The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation. *Language and Speech,* 50, 125–142. doi: 10.1177/00238309070500010601.

Munson, B., & Babel, M. (2007). Loose lips and silver tongues, or, projecting sexual orientation through speech. *Language and Linguistics Compass,* 1, 416–449. doi: 10.1111/j.1749-818x.2007.00028.x.

Munson, B., McDonald, E. C., DeBoe, N. L., & White, A. R. (2006). The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *Journal of Phonetics,* 34, 202–240. doi: 10.1016/j.wocn.2005.05.003.

Pearcey, S. M., Docherty, K. J., & Dabbs, J. M. jr. (1996). Testosterone and sex role identification in lesbian couples. *Physiology & Behavior,* 60, 1033–1035.

Piccolo, F. (2008). Listener-identified phonetic correlates of gay-, lesbian- and straight-sounding speech (Doctoral dissertation, University of Hawaii). Retrieved from https://scholarspace.manoa.hawaii.edu/handle/10125/20680

Pierrehumbert, J. B., Bent, T., Munson, B., Bradlow, A. R., & Bailey, J. M. (2004). The influence of sexual orientation on vowel production (L). *Journal of the Acoustical Society of America,* 116, 1905–1908.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40, 879–891.

Prothro, E. T., & Melikian, L. H. (1955). Studies in stereotypes: V. Familiarity and the kernel of truth hypothesis. *J Soc Psychol, 41*, 3-10. DOI: 10.1080/00224545.1955.9714248

Rendall, D., Vasey, P. L., & McKenzie, J. (2008). The queen's English: An alternative, biosocial hypothesis for the distinctive features of "gay speech". *Archives of Sexual Behavior,* 37, 188–204. doi: 10.1007/s10508-007-9269-x.

Renn, P. (2003). Speech, male sexual orientation, and childhood gender nonconformity (Bachelor thesis, University of Texas, Austin). Retrieved from http://homepage.psy.utexas.edu/homepage/class/Psy158H/PrevHonors/Z111/project.htm

Rieger, G., Linsenmeier, J. A. W., Gygax, L., & Bailey, J. M. (2008). Sexual orientation and childhood gender nonconformity: Evidence from home videos. *Developmental Psychology,* 44, 46–58. doi: 10.1037/0012-1649.44.1.46.

Rieger, G., Linsenmeier, J. A. W., Gygax, L., Garcia, S., & Bailey, J. M. (2010). Dissecting "gaydar": Accuracy and the role of masculinity–femininity. *Archives of Sexual Behavior,* 39, 124–140. doi: 10.1007/s10508-008-9405-2.

Rule, N. O. (2011). The influence of target and perceiver race in the categorisation of male sexual orientation. *Perception,* 40, 830–839.

Rule, N. O., & Ambady, N. (2008). Brief exposures: Male sexual orientation is accurately perceived at 50ms. *Journal of Experimental Social Psychology,* 44, 1100–1105. doi: 10.1016/j.jesp.2007.12.001.

Rule, N. O., Ambady, N., Adams, R. B., & Macrae, C. N. (2008). Accuracy and awareness in the perception and categorization of male sexual orientation. *Journal of Personality and Social Psychology,* 95, 1019–1028. doi: 10.1037/a0013194.

Rule, N. O., Ambady, N., & Hallett, K. C. (2009). Female sexual orientation is perceived accurately, rapidly, and automatically from the face and its features. *Journal of Experimental Social Psychology,* 45, 1245–1251. doi: 10.1016/j.jesp.2009.07.010.

Rule, N. O., Ishii, K., Ambady, N., Rosen, K. S., & Hallett, K. C. (2011). Found in translation: Cross-cultural consensus in the accurate categorization of male sexual orientation. *Personality and Social Psychology Bulletin,* 37, 499–1507.

Rule, N. O., Macrae, C. N., & Ambady, N. (2009). Ambiguous group membership is extracted automatically from faces. *Psychological Science,* 20, 441–443.

Rule, N. O., Rosen, K. S., Slepian, M. L., & Ambady, N. (2011). Mating interest improves women's accuracy in judging male sexual orientation. *Psychological Science,* 22, 881–886. doi: 10.1177/0956797611412394.

Runge, T. E., Frey, D., Gollwitzer, P. E., Helmreich, R. L., & Spence, J. T. (1981). Masculine (instrumental) and feminine (expressive) traits: A comparison between students in the United States and West Germany. *Journal of Cross-Cultural Psychology,* 12, 142–162.

Schubert, T. W., & Otten, S. (2002). Overlap of self, ingroup, and outgroup: Pictorial measures of self-categorization. *Self and Identity,* 1, 353–376.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379–423, 623–656.

Shelp, S. G. (2002). Gaydar: visual detection of sexual orientation among gay and straight men. *Journal of Homosexuality,* 44, 1–14.

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass,* 3, 621–640. DOI: 10.1111/j.1749-818x.2009.00125.x.

Singh, D., Vidaurri, M., Zambarano, R. J., & Dabbs, J. M., jr. (1999). Lesbian erotic role identification: Behavioral, morphological, and hormonal correlates. *Journal of Personality and Social Psychology,* 76, 1035–1049.

Sisson, C. (2003). Phonetic cues in the evaluation of gay male speech in Canadian English and Québec French. Retrieved from http://www.cs.cmu.edu/~./csisso/docs/SocioII-paper.pdf

Skorska, M. N., Geniole, S. N., Vrysen, B. M., McCormick, C. M., & Bogaert, A. F. (2014). Facial structure predicts sexual orientation in both men and women. *Archives of Sexual Behavior.* doi*:* 10.1007/s10508-014-0454-4.

Smyth, R., Jacobs, G., & Rogers, H. (2003). Male voices and perceived sexual orientation: An experimental and theoretical approach. *Language in Society,* 32, 329–350.

Smyth, R., & Rogers, H. (2008). Do gay-sounding men speak like women? *Toronto Working Papers in Linguistics*, 27, 129-144.

Spence, J. T., Helmreich, R. L., & Stapp, J. (1975). Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity. *Journal of Personality and Social Psychology,* 32, 29–39.

Stern, C., West, T. V., Jost, J. T., & Rule, N. O. (2013). The politics of gaydar: Ideological differences in the use of gendered cues in categorizing sexual orientation. *Journal of Personality and Social Psychology,* 104, 520–541.

Sulpizio, S., Fasoli, F., Maass, A., Paladino, M. P., Vespignani, F., Eyssel, F., & Bentler, D. (2015). The sound of voice: Voice-based categorization of speakers' sexual orientation within and across languages. *PLoS ONE,* 10, e0128882. doi:10.1371/journal.pone.0128882

Sylva, D., Rieger, G., Linsenmeier, J. A. W., & Bailey, J. M. (2010). Concealment of sexual orientation. *Archives of Sexual Behavior,* 39, 141–152. doi: 10.1007/s10508-008-9466-2.

Tabak, J. A., & Zayas, V. (2012). The roles of featural and configural face processing in snap judgments of sexual orientation. *PLoS ONE,* 7, S. e36671. doi:10.1371/journal.pone.0036671

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In S. Worchel & W. G. Austin (Eds.), *The social psychology of intergroup relations* (pp. 33 - 47). Monterey, CA: Brooks Cole.

Terango, L. (1966). Pitch and duration characteristics of oral reading. *Journal of Speech & Hearing Research,* 9, 590–595.

Tskhay, K. O., Feriozzo, M. M., & Rule, N. O. (2013). Facial features influence the categorization of female sexual orientation. *Perception,* 42, 1090–1094.

Tskhay, K. O., Krendl, A. C., & Rule, N. O. (2016). Age-related physical changes interfere with judgments of male sexual orientation from faces. *Personality and Social Psychology Bulletin,* 42, 1217–1226. doi: 10.1177/0146167216653585.

Tskhay, K. O., & Rule, N. O. (2013a). Accuracy in categorizing perceptually ambiguous groups: A review and meta-analysis. *Personality and Social Psychology Review,* 17, 72–86. doi: 10.1177/1088868312461308.

Tskhay, K. O., & Rule, N. O. (2013b). Accurate identification of a preference for insertive versus receptive intercourse from static facial cues of gay men. *Archives of Sexual Behavior,* 42, 1217–1222.

Tskhay, K. O., & Rule, N. O. (2015). Emotions facilitate the communication of ambiguous group memberships. *Emotion,* 15, 812–826. doi: 10.1037/emo0000077.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. New York, NY, USA: Basil Blackwell, Inc.

Valentova, J. V., & Havlíček, J. (2013). Perceived sexual orientation based on vocal and facial stimuli is linked to self-rated sexual orientation in Czech men. *PLoS ONE,* 8, e82417. doi: 10.1371/journal.pone.0082417.

Valentova, J. V., Kleisner, K., Havlíček, J., & Neustupa, J. (2014). Shape differences between the faces of homosexual and heterosexual men. *Archives of Sexual Behavior,* 43, 353–361. doi: 10.1007/s10508-013-0194-x.

Valentova, J. V., Rieger, G., Havlíček, J., Linsenmeier, J. A. W., & Bailey, J. M. (2011). Judgments of sexual orientation and masculinity-femininity based on thin slices of behavior: A cross-cultural comparison. *Archives of Sexual Behavior,* 40, 1145–1152.

van Borsel, J., & van de Putte, A. (2014). Lisping and male homosexuality. *Archives of Sexual Behavior,* 43, 1159–1163. doi: 10.1007/s10508-014-0262-x.

van Borsel, J., Vandaele, J., & Corthals, P. (2013). Pitch and pitch variation in lesbian women. *Journal of Voice,* 27, 656.e13. doi: 10.1016/j.jvoice.2013.04.008.

Waksler, R. (2001). Pitch range and women's sexual orientation. *WORD,* 52, 69–77.

Woolery, L. M. (2007). Gaydar: a social-cognitive analysis. *Journal of Homosexuality,* 53, 9–17. doi: 10.1300/J082v53n03_02.

Zimman, L. (2010). Female-to-male transsexuals and gay-sounding voices: A pilot study. *Colorado Research in Linguistics,* 22, 1–21.

Tables

Table 1. Bivariate Correlations of Actual Straightness and Perceived Straightness (Experiment 1) and Perceived Gender-Role Conformity (GRC, Experiment 2) in the Whole Sample (Below Diagonal) and Gender-Separated Sub-Samples (Above Diagonal, Female/Male).

	Actual		Perceived straightness			Perceived GRC	
	Straight-ness	GRC	Voice	Face	Voice + face	Voice	Face
Actual							
Straightness		**.53/.43**	**.38/.41**	**.56/.52**	**.54/.62**	(.31)/**.45**	**.56/.54**
GRC	**.43**		(.31)/**.34**	**.57/.38**	**.55/.47**	**.53**/(.14)	**.56/.43**
Perceived straightness							
Voice	**.35**	**.27**		**.41/.37**	**.47/.63**	**.74/.78**	**.35**/(.21)
Face	**.52**	**.47**	**.35**		**.98/.92**	**.41/.43**	**.88/.72**
Voice + face	**.56**	**.50**	**.54**	**.96**		**.46/.64**	**.89/.70**
Perceived GRC							
Voice	**.39**	**.32**	**.62**	**.40**	**.52**		**.35/.35**
Face	**.54**	**.49**	**.26**	**.81**	**.81**	**.34**	

Note. Correlations for the whole sample are presented below the diagonal (n = 72; for correlations involving perceived gender-role conformity n = 71). *n* = 36 for female and male sub-samples; for correlations of the female sub-sample involving perceived gender-role conformity *n* = 35. Abbreviations: Gender-role conformity (GRC). Actual straightness refers to straight attraction index. Higher scores for actual and perceived straightness indicating a higher degree of straightness. Actual GenCon refers to Traditional Masculinity-Femininity scale. Higher scores for actual and perceived GenCon indicate higher gender-role conformity. All correlations are statistically significant at $\alpha \leq .05$ except for those in parentheses.

Table 2. Regression Analyses of Perceived Straightness in Combination Condition on Perceived Straightness in Face and in Voice Condition for the Whole Sample, Female Sub-Sample, and Male Sub-Sample in Experiment 1.

		Whole sample	Female sub-sample	Male sub-sample
Step 1				
Model-specific	Adj. R^2	.91	.97	.85
	F ($df1$, $df2$)	744.78 (1, 70)	983.73 (1, 34)	195.88 (1, 34)
	p	< .001	< .001	< .001
Face (predictor-specific)	B	1.02	.97	1.12
	SE	.04	.03	.08
	β	.96	.98	.92
	p	< .001	< .001	< .001
Step 2	R^2*-change*	.05	.01	.10
	F ($df1$, $df2$)	84.23 (1, 69)	7.16 (1, 33)	59.32 (1, 33)
	p	< .001	.011	< .001
Model-specific	Adj. R^2	.96	.97	.94
	F ($df1$, $df2$)	857.28 (2, 69)	584.61 (2, 33)	295.59 (2, 33)
	p	< .001	< .001	< .001
Face (predictor-specific)	B	.94	.93	.97
	SE	.03	.03	.05
	β	.88	.95	.80
	p	< .001	< .001	< .001
Voice (predictor-specific)	B	.40	.29	.44
	SE	.04	.11	.06
	β	.23	.09	.33
	p	< .001	.011	< .001

Table 3. Bivariate Correlations of Acoustic Characteristics and Perceived Straightness (Experiment 1) in Voice, Face, and Voice+Face Condition, and Perceived Gender-Role Conformity (Experiment 2) in Voice and Face Condition for Male (M) and Female (F) Targets.

	Perceived straightness in…				Perceived gender-role conformity in voice	
	voice		voice+face			
	M[1]	F[1]	M[1]	F[1]	M[1]	F[2]
Vowel space characteristics						
Vowel space expansion	(-.04)	**.37**	(.07)	**.40**	(.05)	**.47**
Mean F1	(.17)	**.57**	(.06)	(.31)	(-.05)	**.54**
Mean F2	(-.22)	**.48**	(-.24)	(.20)	(-.26)	**.62**
/aː/ F1	(.06)	**.63**	(-.01)	**.51**	(-.11)	**.55**
/aː/ F2	(-.22)	**.34**	(-.20)	**.36**	**-.38**	**.52**
/ɪ/ F1	(-.15)	(-.01)	(-.22)	(-.18)	(-.17)	(.07)
/ɪ/ F2	(-.11)	**.42**	(-.08)	(.09)	(.00)	**.56**
/ɔ/ F1	**.37**	**.35**	(.30)	(.05)	(.16)	(.32)
/ɔ/ F2	(-.06)	(.30)	(-.14)	(-.06)	(-.08)	(.26)
Fundamental frequency characteristics						
f0 mean	(-.24)	**.37**	(-.07)	(.26)	**-.50**	**.47**
f0 *SD*	(.21)	(-.33)	(.29)	(-.02)	(.27)	(-.05)
f0 2.5th percentile	**-.47**	**.35**	(-.32)	(.05)	**-.64**	(.14)
f0 97.5th percentile	(.13)	(-.05)	(.25)	(.09)	(.22)	(.02)

Note. All correlations are statistically significant at α ≤ .05 except for those in parentheses. Abbreviations: F1 (first formant), F2 (second formant), and f0 (fundamental frequency).

[1] n = 36. [2] n = 35.

Table 4. Bivariate Correlations of Psychological Characteristics and Perceived Straightness (Experiment 1) in Voice, Face, and Voice+Face Condition, and Perceived Gender-Role Conformity (Experiment 2) in Voice and Face Condition for the Overall Sample (All), the Male (M) and the Female Sub-Sample (F).

	Perceived straightness in…									Perceived gender-role conformity in…					
	… voice			… face			… voice+face			… voice			… face		
	All[1]	M[2]	F[2]	All[1]	M[2]	F[2]	All[1]	M[2]	F[2]	All[3]	M[2]	F[4]	All[3]	M[2]	F[4]
Actual gender-role conformity															
TMF	**.27**	**.34**	(.31)	**.47**	**.38**	**.57**	**.50**	**.47**	**.55**	**.32**	(.14)	**.53**	**.49**	**.43**	**.56**
GEPAQ-F	**.44**	(.19)	(.05)	(.21)	**.45**	(.09)	**.28**	**.49**	(.11)	(.00)	(.18)	(-.00)	**.24**	**.52**	(.11)
GRB-F	(.23)	(.04)	(.20)	**.37**	**.47**	(.29)	**.39**	**.42**	(.31)	**.24**	(.23)	**.31**	**.41**	**.37**	**.42**
Childhood Gender Role Conformity	(.14)	(.28)	(.29)	**.32**	**.50**	(.27)	**.35**	**.58**	(.27)	**.49**	**.38**	**.59**	**.33**	**.49**	(.27)
Social group affiliation															
Self - Lesbians/gay men	(.14)	(.16)	(.19)	**.46**	**.52**	**.44**	**.46**	**.51**	**.44**	**.25**	(.28)	(.23)	**.50**	**.52**	**.48**
Self - Straight women/men	**.37**	(.29)	**.42**	**.48**	**.36**	**.59**	**.49**	**.39**	**.57**	**.37**	**.43**	**.35**	**.52**	**.50**	**.53**
Social environment															
Contact to boys during childhood	(-.19)	**.37**	(.15)	(.05)	**.50**	(.07)	(.03)	**.55**	(.07)	(.22)	(.32)	(.31)	(.09)	**.46**	(.17)
Contact to girls during childhood	**.37**	(.03)	(.10)	(.17)	(.19)	(.08)	(.22)	(.19)	(.07)	(.09)	(.19)	(.21)	(.19)	(.22)	(.07)

Gender distribution of friends	(.14)	(-.07)	(.19)	(.06)	**.39**	(-.23)	(.10)	**.34**	(-.20)	(.22)	(.22)	**.26**	(.15)	**.37**	(-.12)
Sexual orientation of same-gender friends	**.30**	**.36**	**.38**	**.40**	**.47**	**.39**	**.44**	**.56**	**.38**	**.32**	(.26)	**.39**	**.38**	**.50**	**.27**
Contact to lesbians/gay men	**.33**	**.41**	(.28)	**.29**	**.34**	(.27)	**.33**	**.41**	(.26)	**.26**	(.25)	**.37**	**.31**	(.28)	(.24)
Contact to straight women/men	(.16)	(.04)	**.34**	**.31**	**.46**	(.19)	**.34**	**.43**	(.23)	**.33**	(.00)	(.16)	(.06)	**.50**	(.12)

Note. For all psychological characteristics, positive scores indicate gender- and/or straightness-conform characteristics and negative scores vice versa. All psychological characteristics with slashes referred to female groups for female participants and male groups for male participants. All correlations are statistically significant at $\alpha \leq .05$ except for those in parentheses. Traditional Masculinity-Femininity (TMF), Femininity scale of the German Extended Personality Attributes Questionnaire (GEPAQ-F), Femininity scale of the Gender-Role Behavior (GRB-F), and Childhood Gender-Role Behavior (CGRC).

[1] $n = 72$. [2] $n = 36$. [3] $n = 71$. [3] $n = 35$.

Figures

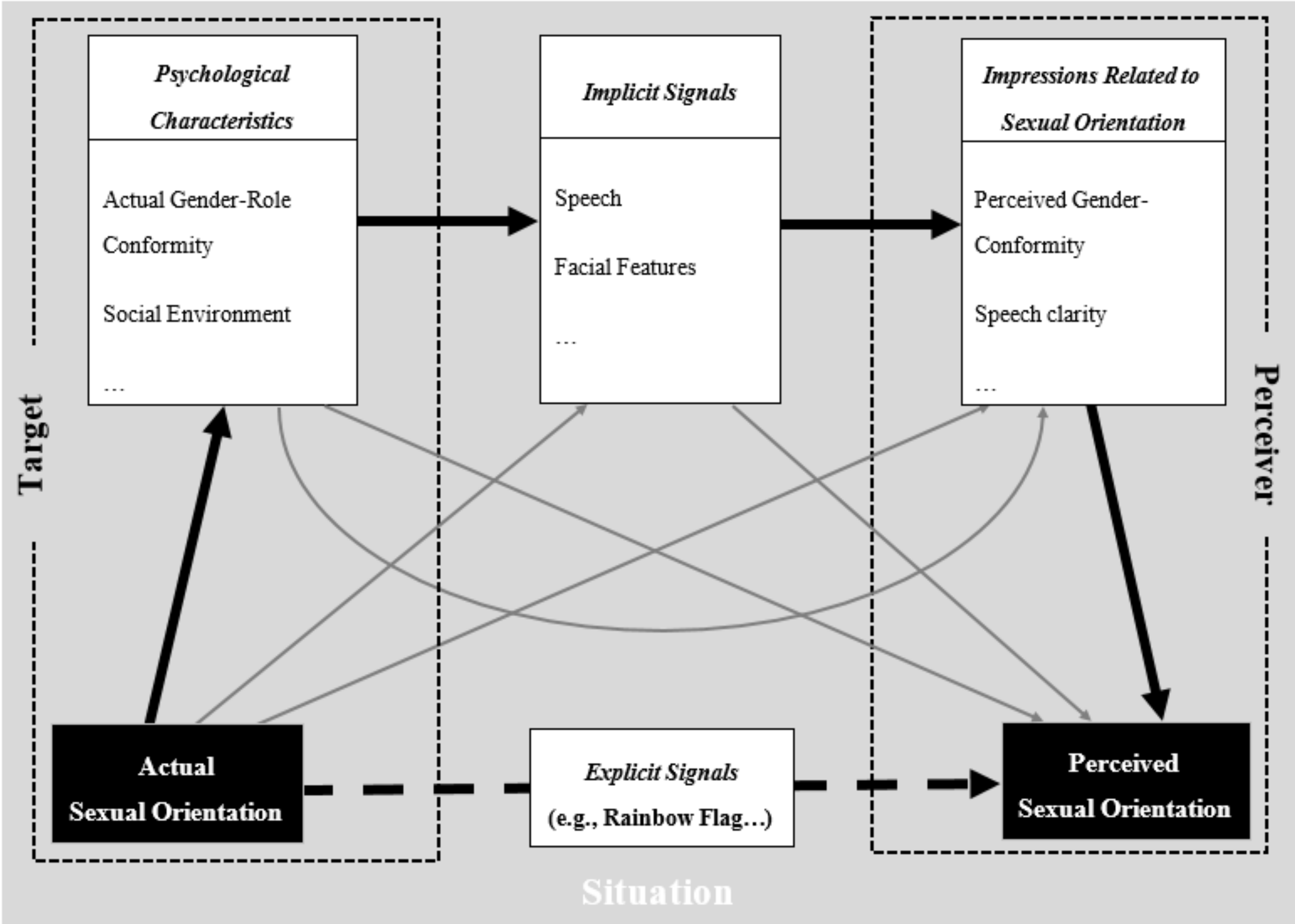


Figure 1. Expression and Perception of Sexual Orientation Model (EPSOM).

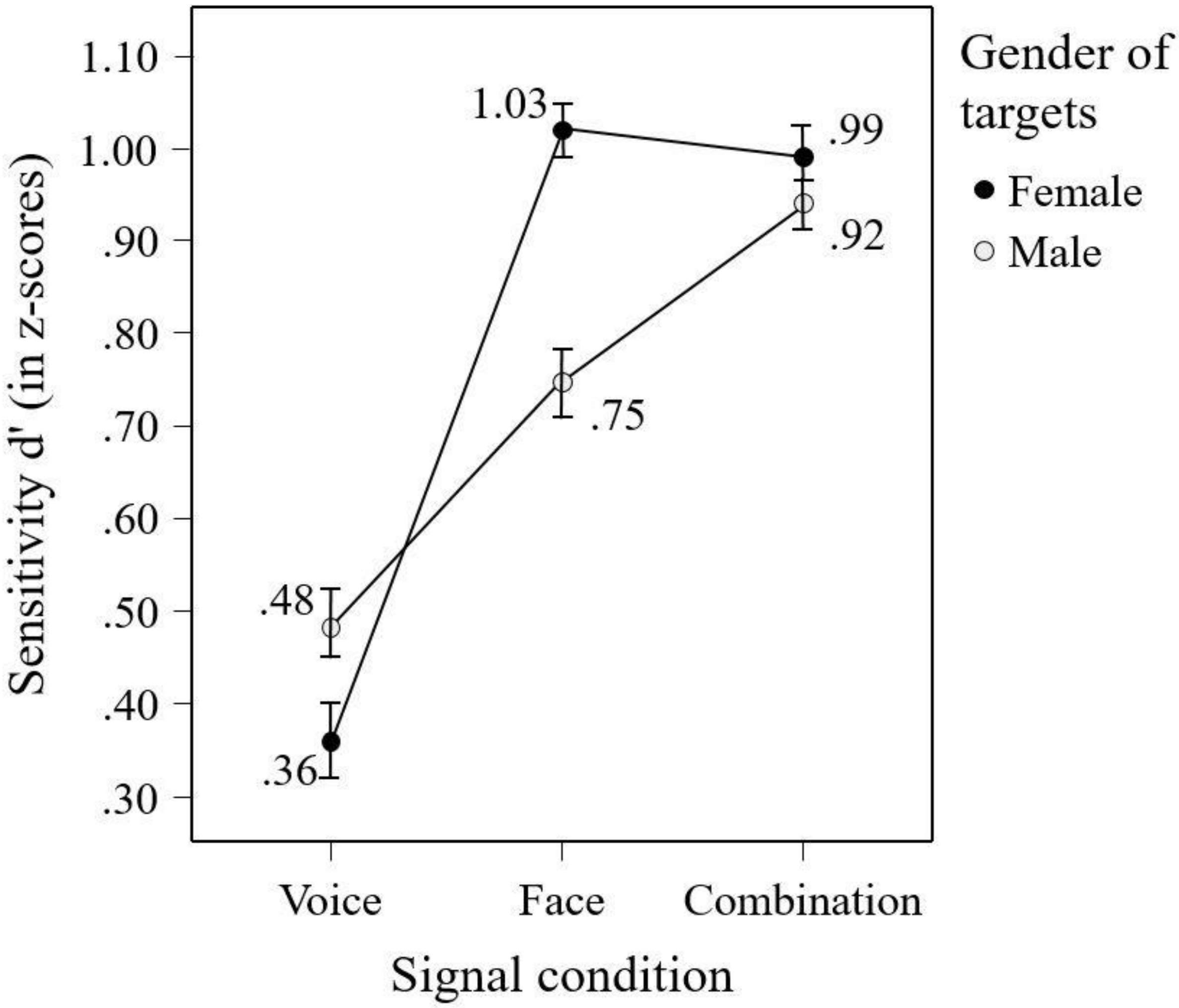


Figure 2. Means for Sensitivity d’ for each Signal Condition by Gender. More Positive d’ Scores Indicate more Correct Classification of Targets’ Sexual Orientation. Error Bars Show Standard Errors of Means.

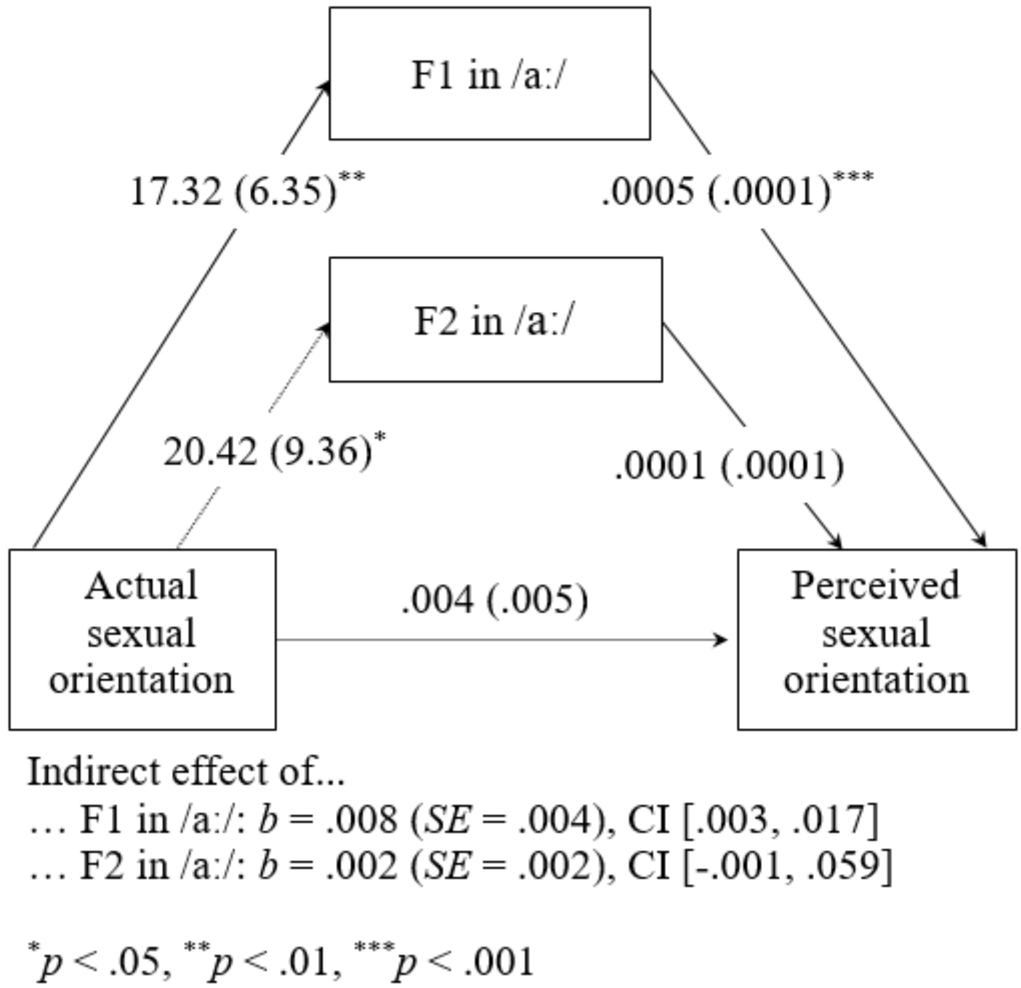


Figure 3. Mediation of the Effect of Actual on Perceived Sexual Orientation by F1 and F2 in /aː/ for Female Targets in Voice Condition.

Manuscript 5 – Related Appendix

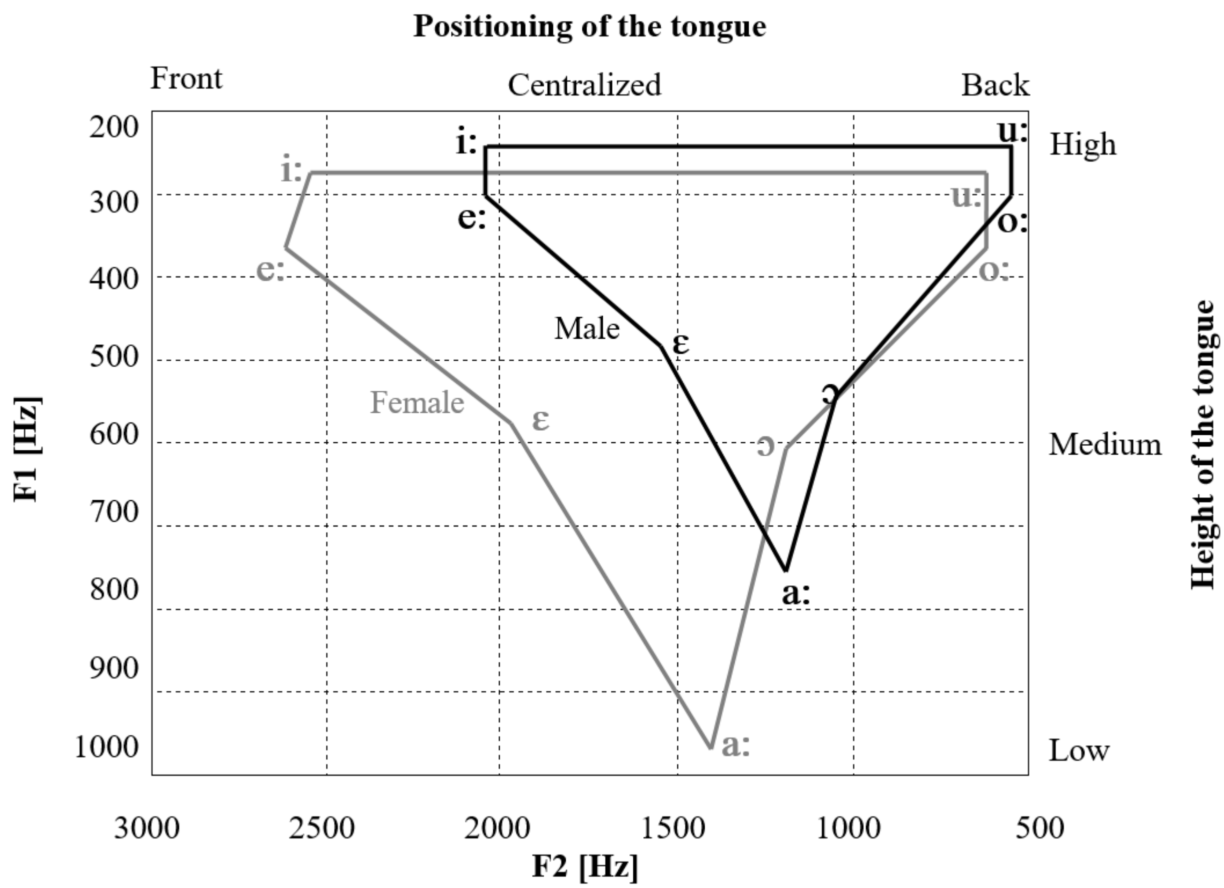


Figure A1. Gender Differences in Vowel Spaces between Female and Male German Speakers (adapted from Simpson & Ericsdotter, 2007).

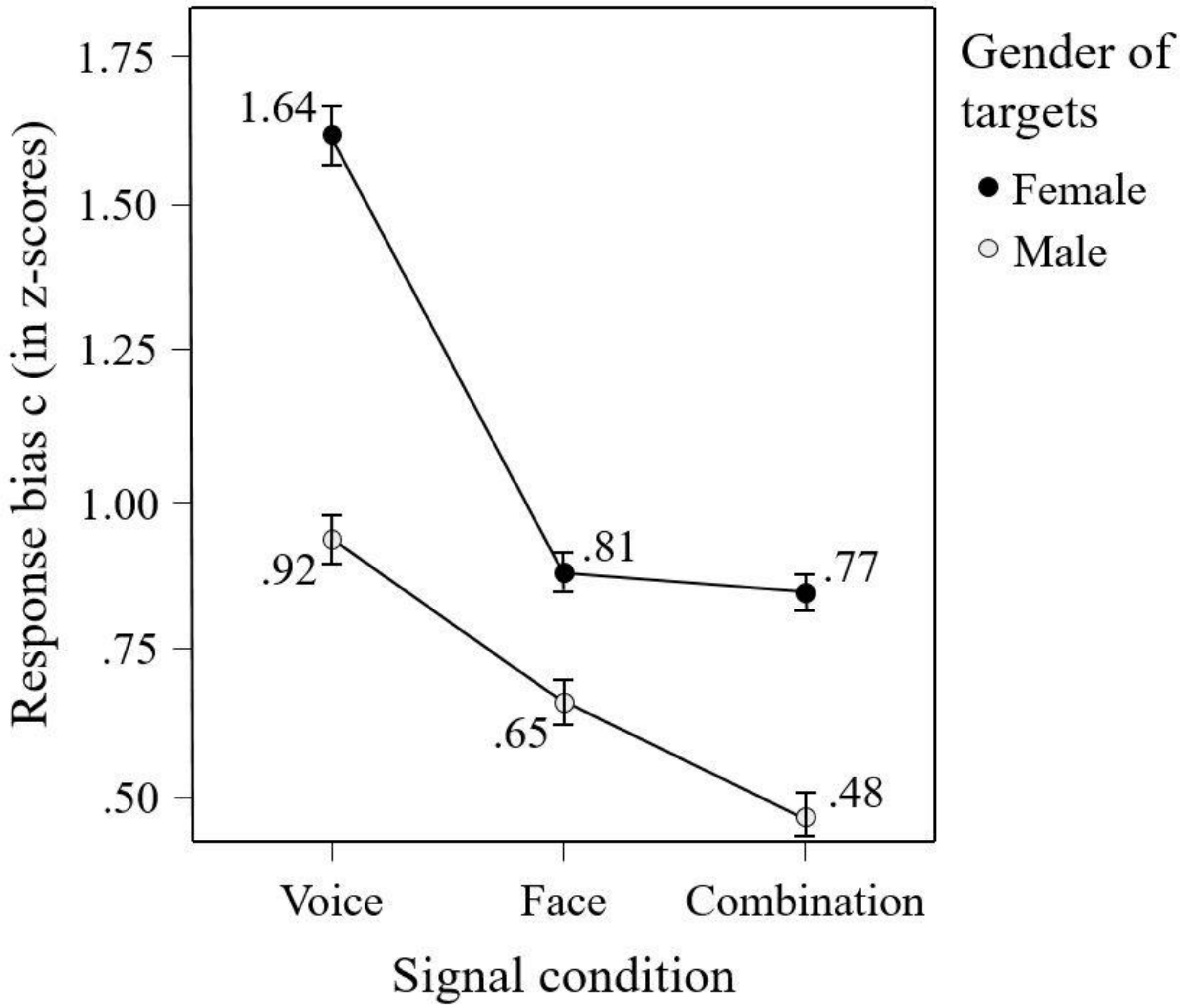


Figure A2. Means for Response Bias C for each Signal Condition by Gender. More Positive C Scores Indicate a Stronger Tendency for Straight Responses. Error Bars Show Standard Errors of Means.

Response bias analysis

Results for the response bias C are illustrated in Figure 3 and indicate that in every condition there was a strong bias to respond straight. There was an interaction, $F(2, 99) = 15.44$, $p < .001$, $\eta^2_p = .24$, as well as main effect of signal condition, $F(2, 99) = 83.53$, $p < .001$, $\eta^2_p = .63$, and targets' gender, $F(1, 100) = 49.69$, $p < .001$, $\eta^2_p = .33$. Simple-effects tests with Bonferroni adjustment revealed that for male targets the tendency to judge targets as straight decreased from voice over face to combination condition, $F(2, 99) = 30.44$, $p < .001$, $\eta^2_p = .38$ (all pair-wise *p*s $\leq .002$). For female targets the bias to respond straight was stronger in voice condition compared to face and combination condition, $F(2, 99) = 72.04$, $p < .001$, $\eta^2_p = .59$ (both pair-wise *p*s $\leq .001$), but there was no difference between face and combination condition (pair-wise $p = .237$). However, the tendency to judge targets as straight was significantly stronger for female targets in each signal condition compared to male targets, all $F(1, 100) \geq 4.68$, all $p \leq .033$, all $\eta^2_p \geq .05$ (all pair-wise *p*s $\leq .033$). Hence, results on response bias between signals separated for female and male targets mirrored findings for accuracy of implicit signals proposed by the EPSOM. While accuracies for male targets increased from voice over face to combination condition and bias to respond straight decreased, for female targets higher accuracies in face-including conditions than in voice condition were accompanied by a lower straight response bias.

Correlations on Gender-Role Conformity

Perceived Gender-Role Conformity in Different Conditions

Could judgments about gender-role conformity based on one signal be used to deduce judgments based on the other signal? The more gender conforming targets were judged based on voices, the more gender conforming they were judged based on faces ($r = .34$, $p = .003$). There was no difference for female and male targets ($z = -.01$, $p = .496$).

Cross-dimensional Correlations

Is actual sexual orientation (or gender-role conformity) linked to perceived gender-role conformity (or sexual orientation)? This is indeed the case (see Table 1): The straighter targets described themselves, the more likely they were perceived as gender conforming based on faces ($r = .54$, $p < .001$) and voices ($r = .39$, $p = .001$); the more gender conforming targets described themselves the straighter they were perceived based on faces ($r = .47$, $p < .001$) and voices ($r = .27$, $p = .022$; $z = -1.62$, $p = .053$). Hence, actual sexual orientation was reflected in perceived gender-role conformity and actual gender-role conformity was reflected in perceived sexual orientation.

Curriculum Vitae

Sven Kachel

Universität Koblenz-Landau

FB 8: Sozial-, Umwelt- und Wirtschaftspsychologie
Fortstraße 7 | 76829 Landau | Germany
Arbeitsanschrift

Email: sven.kachel@uni-jena.de

Born on 20.03.1987 in Jena (Germany).

Education

09/2016-now	Psychotherapist training with cognitive behavioral focus at the IVT Dresden (Germany)
02/2013	Bachelor of Arts in Communication and Speech Sciences, Friedrich Schiller University of Jena, Germany (1.5) B.A. Thesis „Mediale Konstruktion von Homosexualität” (Medial Construction of Homosexuality) (1.3)
09/2012	Diploma in Psychology (equivalent to M.Sc.); Friedrich-Schiller-University of Jena, Germany (1.9) Diploma Thesis: „Geschlechterrolle oder sexuelle Identität? Eine phonetisch-psychologische Untersuchung weiblicher Sprechweisen” [Gender role or sexual identity? A phonetic-psychological analysis of female speech] (1.0)
2008-2013	Academic studies of Communication Sciences (major subject) and Speech Sciences (minor subject) at the Friedrich Schiller University Jena
2005-2012	Academic studies of Psychology at the Friedrich Schiller University Jena (minor subject: German Linguistics)
06/2005	Abitur (1.4)
1997-2005	Secondary school “Ernst-Abbe-Gymnasium” Jena

1993-1997	Primary school “Schule an der Trießnitz” Jena

Positions

1/2013-now	PhD student and research assistant in the DFG Research Unit Person Perception (Friedrich Schiller University of Jena, Germany) and at the Department of Social, Environmental, and Economic Psychology (University Koblenz-Landau, Germany)
10/2009-10/2011	Free lancing at radio station Funkwerk in Erfurt: Presentation of the kids radio programme Fledermausfunk and the electronic music radio programme Frequenzmodulation
12/2008-01/2011	Internship at Podsports (activities in voice coaching and compilation of auditory content on physical trainings)
01/2007-09/2009	Research internship and student assistant in the DFG-funded research project “Brain activation during automatic and controlled stimulus processing before and after psychotherapeutic treatment on specific phobia” at the Department of Clinical and Biological Psychology, Friedrich Schiller University Jena
04/2006-10/2006	Student assistant at the Thuringian University and Federal State Library

Publications

Papers in Peer-Reviewed Journals

Kachel, S., Simpson, A.P., & Steffens, M.C. (2017). Acoustic correlates of sexual orientation and gender-role self-concept in women’s speech. *Journal of the Acoustical Society of America,* 141, 4793-4809.

Kachel, S., Steffens, M.C., & Niedlich, C. (2016). Traditional masculinity and femininity: Validation of a new scale assessing gender roles. *Frontiers in Psychology*, 7:956. DOI: 10.3389/fpsyg.2016.00956

Book Chapters

Steffens, M.C., Niedlich, C., **Kachel, S.**, & Methner, N. (2016). Impression formation of applicants differing in sexual orientation: An attempt to integrate theoretical models and a review of the empirical evidence. In F. Earley (Ed.), *Sexual orientation: Perceptions, discrimination and acceptance.* New York: Nova Science Publishers.

Selected Conference Presentations

Kachel, S., Radtke, A., Skuk, V.G., Zäske, R., Simpson, A.P., & Steffens, M.C. (2016). *Untersuchung der sexuellen Orientierung mittels Voice Morphing.* [Examining sexual orientation using voice morphing] Talk at the 50th Conference of the German Society for Psychology, 18.-22.09.2016, Leipzig, Germany.

Kachel, S., Steffens, M.C., & Simpson, A.P. (2015). *'Testing your Gaydar!' – Do mode and target characteristics influence straight judgements?* Talk at the 15th Conference of the German Society for Psychology's Section Social Psychology, 06.-09.09.2015, University of Potsdam, Germany.

Simpson, A.P., **Kachel, S.**, & Steffens, M.C. (2014). *Is straight vs. GLB too coarse? Investigating the phonetic correlates of sexual orientation in German women.* Poster presented at the BAAP Colloquium, 07.-09.04.2014, Oxford, Great Britain.

Kachel, S., Simpson, A.P., & Steffens, M.C. (2013). *Lesbians do not speak differently than straight women, but they do speak differently than lesbians.* Poster presented at the Final Conference of the Marie Curie Initial Training Network Language, Cognition and Gender, 13.-16.06.2013, Bern, Switzerland.

Kachel, S., Simpson, A.P., & Steffens, M.C. (2013). *Phonetic Parametrization of Sexual Orientation and Gender in German.* Poster presented at the 55th Conference of Experimental Psychologists, 24-27.03.2013, Vienna, Austria.

Kachel, S. (2013). *Geschlechterrolle oder sexuelle Identität? Eine phonetisch-psychologische Untersuchung weiblicher Sprechweisen.* [Gender role or sexual identity? A phonetic-psychological analysis of female speech] Invited talk at the 14th Conference of the German Society of Psychology's Section Social Psychology, 01.-04.09.2013, Fernuniversity Hagen, Germany.

Kachel, S., Simpson, A.P., & Steffens, M.C. (2012). *Akustische Korrelate von Gender und sexueller Orientierung – Perzeptive Hinweisreize der Sexualität in der menschlichen Stimme.* [Acoustic correlates of gender and sexual orientation – Perceptual cues of sexuality in the humen voice] Poster presented at the 8th Phonetik & Phonologie, 12.-13.10.2012, Jena, Germany.

Simpson, A.P., **Kachel, S.**, & Steffens, M.C. (2012). *Acoustic correlates of gender and sexual orientation – Perceptual cues to sexuality in the human voice.* Talk at the 6th Workshop of the Person Perception Research Unit „Person Perception: Individual differences and differences between groups", 10.-11.10.2012, Jena, Germany.

Eidesstattliche Erklärung

Hiermit erkläre ich, Sven Kachel, dass ich die Synopse der vorliegenden Dissertation selbstständig verfasst habe und keine anderen als die angegebenen Quellen genutzt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken habe ich als solche kenntlich gemacht.

Für die gemeinsam verfassten Publikationen habe ich folgende individuelle Beiträge erbracht.

Manuscript 1. Kachel, S., Steffens, M. C., & Niedlich, C. (2016). Traditional Masculinity and Femininity: Validation of a New Scale Assessing Gender Roles. *Frontiers in Psychology,* 7:956. DOI: 10.3389/fpsyg.2016.00956.

- Beteiligung an der Konzeption des Papers und an der Überarbeitung nach Vorlage der Reviewer*innengutachten. Die Skala wurde von Melanie C. Steffens und Kornelia Schertzl entwickelt.
- *Einleitung:* Mithilfe bei der Literaturrecherche. Beteiligung an der Dokumentation.
- *Pilotstudie:* Maßgeblich verantwortlich für die Konzeption und Dokumentation. Allein verantwortlich für die Auswertung.
- *Studie 2:* Maßgeblich verantwortlich für Konzeption, für die Datenauswertung sowie für die Dokumentation.
- *Diskussion:* Mithilfe bei der Erstellung.

Manuscript 2. Kachel, S., Simpson, A. P., & Steffens, M. C. (2017). Acoustic Correlates of Sexual Orientation and Gender-Role Self-Concept in Women's Speech. *Journal of the Acoustical Society of America,* 141, 4793-4809. DOI: 10.1121/1.4988684

- Maßgeblich verantwortlich für die Konzeption des Papers und für die Überarbeitung nach Vorlage der Reviewer*innengutachten.
- *Einleitung:* Maßgeblich verantwortlich für die Literaturrecherche und die Dokumentation.
- *Studie 1:* Maßgeblich verantwortlich für die Konzeption, Auswertung und Dokumentation. Allein verantwortlich für die Datenerhebung und –vorverarbeitung. Die Daten wurden im Rahmen meiner Diplomarbeit gesammelt. Für die Dissertationsschrift wurden die Daten aus Studie 1 restrukturiert und nach veränderten methodischen Gesichtspunkten ausgewertet. Zudem wurde ein klarer inhaltlicher Schwerpunkt im Zusammenhang mit dem Thema des Manuskripts gesetzt.
- *Studie 2:* Maßgeblich verantwortlich für Konzeption, für die Datenauswertung sowie für die Dokumentation. Mithilfe bei der Vorverarbeitung der Daten.
- *Diskussion:* Maßgeblich verantwortlich für die Erstellung.

Manuscript 3. Kachel, S., Simpson, A. P., & Steffens, M. C. (2017). Do I Sound Straight? – Acoustic Correlates of Actual and Perceived Sexual Orientation and Masculinity/Femininity in Men's Speech. In Press in *Journal of Speech, Language, and Hearing Research.*

- Maßgeblich verantwortlich für die Konzeption des Papers und für die Überarbeitung nach Vorlage der Reviewer*innengutachten.
- *Einleitung:* Maßgeblich verantwortlich für die Literaturrecherche und die Dokumentation.
- *Studie:* Maßgeblich verantwortlich für die Konzeption, Auswertung und Dokumentation. Mithilfe bei der Datenvorverarbeitung.
- *Diskussion:* Maßgeblich verantwortlich für die Erstellung.

Manuscript 4. Kachel, S., Radtke, A., Skuk, V. G., Zäske, R., Simpson, A. P., & Steffens, M. C. (2017). Do They All Speak the Same? – Investigating Sexual Orientation Information Using Voice Averages. Invited to Revise and Resubmit in *PLOS ONE*.

- Maßgeblich verantwortlich für die Konzeption des Papers und für die Überarbeitung nach Vorlage der Reviewer*innengutachten.
- *Einleitung:* Maßgeblich verantwortlich für die Literaturrecherche und die Dokumentation.
- *Studie:* Maßgeblich verantwortlich für die Konzeption, für die Auswertung und für die Dokumentation.
- *Diskussion:* Maßgeblich verantwortlich für die Erstellung.

Manuscript 5. Kachel, S., Steffens, M. C., & Simpson, A. P. (2017). The Expression and Perception of Sexual Orientation Model: Speech Based Evidence. Submitted to *Journal of Personality and Social Psychology.*

- Maßgeblich verantwortlich für die Konzeption des Papers.
- *Einleitung:* Maßgeblich verantwortlich für die Literaturrecherche und die Dokumentation.
- *Studie 1:* Mithilfe bei der Konzeption der Studie. Maßgeblich verantwortlich für die Auswertung und Dokumentation der Studie.
- *Studie 2:* Mithilfe bei der Konzeption der Studie. Maßgeblich verantwortlich für die Auswertung und Dokumentation der Studie.
- *Diskussion:* Maßgeblich verantwortlich für die Erstellung.

Diese Arbeit habe ich weder in gleicher noch ähnlicher Form einer staatlichen oder anderen Prüfungsbehörde vorgelegt. Jene Teilbeiträge, welche im Rahmen meiner Diplomarbeit bereits einer Prüfungsbehörde vorgelegen haben, wurden für die entsprechenden Publikationen transparent gemacht und vollständig benannt.

Dresden, den 11.07.2017 (Dipl.-Psych. Sven Kachel)