



UNIVERSITÄT
KOBLENZ · LANDAU

Fachbereich 4: Informatik

Entwicklung einer interaktiven 3D-Visualisierung der oro-pharyngo-laryngealen Region für die Sprechtherapie

Masterarbeit
zur Erlangung des Grades
MASTER OF SCIENCE
im Studiengang Computervisualistik

vorgelegt von

Laura Haraké

Erstgutachter: Prof. Dr. Stefan Müller, Institut für Computervisualistik,
Fachbereich Informatik, Universität Koblenz-Landau

Zweitgutachter: Anna Katharina Hebborn, M.Sc., Institut für Computervisualistik,
Fachbereich Informatik, Universität Koblenz-Landau

Koblenz, im November 2014

Erklärung

Ich versichere, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Die Vereinbarung der Arbeitsgruppe für Studien- und Abschlussarbeiten habe ich gelesen und anerkannt, insbesondere die Regelung des Nutzungsrechts.

Mit der Einstellung dieser Arbeit in die Bibliothek bin ich einverstanden. ja nein

Der Veröffentlichung dieser Arbeit im Internet stimme ich zu. ja nein

Koblenz, den 14. November 2014

Kurzfassung

In der vorliegenden Arbeit wird eine prototypische Anwendung für die Logopädie entworfen, die als therapeutisches Werkzeug dient, mit dem die Artikulation deutscher Laute und der Schluckvorgang simuliert werden können. Besonderer Fokus liegt dabei auf der dreidimensionalen Visualisierung der anatomischen Modelle der oro-pharyngolaryngealen Region, die vom Logopäden interaktiv bedient werden kann. Unter Berücksichtigung therapeutischer Anforderungen wird die Frage untersucht, ob eine solche Anwendung im Therapie-Alltag eingesetzt werden und einen Mehrwert zu den üblichen Hilfsmitteln der Sprechtherapie bieten kann.

Abstract

This thesis presents a prototypical application for speech therapy as a therapeutical tool, simulating the articulation of german phonemes and the swallowing reflex. A special attention is given to a three-dimensional visualization of anatomical models of the pharyngolaryngeal region, which can be used in an interactive way. For examining the benefits of such an application in relation to conventional therapy mediums, the thesis considers iteratively the requirements of speech therapists.

Danksagung

Ich bedanke mich bei Frau Dorota Bełtkiewicz, die den Grundstein für die Idee dieser Arbeit legte und inhaltliche Hilfestellungen dazu gab. Mein besonderer Dank gilt Frau Dr. phil. Patricia Sandrieser, in deren logopädischen Fachrunde ich die Anwendung präsentieren konnte. Für die zahlreichen konstruktiven Hilfestellungen möchte ich mich auch bei Frau Eva Spelberg bedanken, in deren Einrichtung ich hospitieren durfte, und bei allen anderen Logopäden, die diese Arbeit freundlicherweise getestet und beurteilt haben: Petra Senger und ihr Team der BDH-Klinik Vallendar, Angelika Reinhardt sowie ihren Kollegen der Pusteblume Fulda und Benedikt Feistel des Therapieinstituts Feistel und Töller in Koblenz.

Inhaltsverzeichnis

1	Einleitung	17
1.1	Motivation	18
1.2	Zielsetzung	20
1.3	Aufbau der Arbeit	20
2	Grundlagen der Anatomie und Lautlehre	23
2.1	Physiologische Sicht	23
2.1.1	Schluckvorgang	26
2.1.2	Artikulation	26
2.2	Einteilung von Sprachlauten	28
3	Grundlagen der Sprachheilkunde	31
3.1	Störungsbilder	32
3.1.1	Störungen der Sprach- bzw. Lautsprachentwicklung	32
3.1.2	Zentrale Sprach- und Sprechstörungen	34
3.1.3	Rhinophonien	35
3.1.4	Schluckstörungen	35
3.2	Therapie	35
3.2.1	Diagnoseansätze	37
3.2.2	Therapiekonzepte	38
4	Verwandte Forschungsansätze	43
5	Konzept	47
5.1	Beschreibung	47
5.2	Anforderungen	49
5.2.1	Therapeutische Anforderungen	49
5.2.2	Funktionale Anforderungen	50
5.3	Herangehensweise	50
5.3.1	Visualisierung	51
5.3.2	Modelle	52

6	Implementierung	55
6.1	Umsetzung des ersten Prototypen	55
6.1.1	Zunge bewegen	57
6.1.2	Ein- und Ausblenden von Objekten	59
6.1.3	Mikrofon-Anbindung	60
6.2	Beurteilung des ersten Prototypen	61
6.3	Umsetzung des finalen Prototypen	64
6.3.1	Zunge bewegen	67
6.3.2	Schluckvorgang	73
6.3.3	Mikrofon-Anbindung	73
6.3.4	Render-Funktion	74
7	Evaluation und Bewertung	75
7.1	Evaluation	75
7.2	Bewertung	77
8	Fazit	83
9	Ausblick	85
A		87
A.1	Fragebogen zum ersten Prototypen	88
A.2	Aussprache deutscher Laute	93

Tabellenverzeichnis

1.1	Ausgangsdaten der Umfrage	19
3.1	Störungsbilder der Komponenten Sprache und Sprechen, Auszug aus Grohnfeldt [Gro12, Tab. 7, S. 74]. Schluckstörungen stellen ein eigenes Störungsbild dar und sind hier nicht aufgeführt	36
6.1	Fragebogen: Welche vorgestellten Funktionen wären Ihnen praktisch wichtig, auf welche könnten Sie verzichten? Mehrfach-Antworten waren hier möglich	62

Abbildungsverzeichnis

2.1	Ansicht von vorne auf die unteren und oberen Zungenbeinmuskeln nach Schindelmeiser [Sch10, S. 85, 86]	24
2.2	Seitliche Ansicht auf die äußeren Muskeln der Zunge nach Schindelmeiser [Sch10, S. 132, 136] und den oro-pharyngo-laryngealen Bereich nach Bartolome [BSM14, S. 13]	25
2.3	Resonanzraum (rot und blau) bei Bildung des Vokals [a] und [i] nach Schindelmeiser [Sch10, S. 138]	27
2.4	Die Unterteilung von Vokalen und Konsonanten im Deutschen über Zungenhöhe und Zungenhebung bzw. Artikulationsort und Artikulationsart nach [WZ11, S. 4, 5]	28
3.1	Aufgabengebiete der Sprachtherapie nach Grohnfeldt [Gro14, S. 17] . .	32
3.2	Bildmaterialien für myofunktionelle Übungen	39
5.1	Ein erstes Mock-Up der Anwendung. Die Bedienoberfläche bietet zwei Ansichten des 3D-Modells, eine Auswahlliste von Lauten, Schieberegler um die Zunge nach ihren Muskeln zu bewegen und Kontrollkästchen um Modellteile ein- und ausblenden zu lassen. Der überwiegende Teil der Bedienoberfläche wurde für den ersten Prototypen übernommen. . .	51
6.1	Zungenmodell mit SimpleDeform-Modifier und der Zungenspitze als Vertex-Gruppe. (a) Der SimpleDeform-Modifier wird um 30 Grad gekrümmt (blaues Gitternetz), (b) Die ausgewählte Vertex-Gruppe (orangefarbenes Gitternetz) schmälert die Zungenspitze, (c) Angelegte Vertex-Gruppe für Zungenkörper (orange) und Seitenränder (blau)	56
6.2	(a), (b) Vollständiges Modell im Ruhezustand, (c) Übersicht über das User-Interface, das auf drei Bereiche reduziert wurde. Beim ersten Prototypen findet sich Funktionalität im Header (1) und im rechten Panel (2). Das Modell lässt sich im Hauptfenster (3) drehen und bewegen. . .	57
6.3	Menü-Bereiche des UI-Panels im Detail. (a) Laute vorführen lassen, (b) Die aufgeklappte Phon-Liste	58

6.4	Menü-Bereich um die Zunge entsprechend der Muskeln zu bewegen . . .	59
6.5	Der Muskel Hyoglossus (grün) lässt die Zunge links und rechts um die lokale X-Achse kippen	60
6.6	Das Anzeige-Panel erlaubt es, Modellteile ein- und ausblenden zu lassen.	60
6.7	Bedienoberfläche und Modell des finalen Prototypen. Ansicht auf das Gesamt-Modell	65
6.8	Die beiden Polygonnetze des Zungenmodells vom ersten und finalen Prototypen im Vergleich. (a) erster Prototyp, (b) finaler Prototyp	66
6.9	Zwei Vertex-Gruppen dienen dem verfeinerten Transformieren der Zungenspitze. (a) Vertex-Gruppe auf dem Polygonnetz des MeshDeform-Modifiers, (b) Vertex-Gruppe auf dem Polygonnetz der Zunge	66
6.10	Liste auszuwählender Phone	68
6.11	Liste auszuwählender Geschwindigkeiten	69
6.12	(a) Liste auszuwählender Tonausgaben. In der Liste steht das x für den ausgewählten Laut und das a für den koartikulierten Laut, (b) Stellung der Zunge während der Zungenruhelage. Der Kiefer ist in dieser Ansicht transparent	70
6.13	Konstanter, konstant nasaler, leichter und kräftiger Phonatationsstrom im Vergleich. (a) konstant beim Laut [o], (b) nasal beim Laut [m], (c) sanft beim Laut [b], (d) kräftig beim Laut [t]	71
6.14	(a) Bedienelemente um die Zunge frei zu bewegen, (b) Skizze des Kollisionstests von der Seite. Mithilfe des Skalarprodukts wird der Vertex der Zunge v_{tongue} auf die Ebene des Flächenenelements projiziert	72
6.15	Die Kollisionserkennung findet während der Bewegung statt. (a) Die Zunge wird nach oben und links verschoben, die Zungenspitze wird nach oben an die Zähne bewegt, (b) Kollisionsmodelle der oberen und unteren Zähne (blau)	73
6.16	Die Schluckbewegungen werden in vier Phasen eingeteilt. Der Bolus ist dunkelblau dargestellt, die Epiglottis ist lila dargestellt. (a) Vorbereitungsphase: Der Bolus liegt in der Zungenschüssel, (b) Orale Phase: Das Velum ist gehoben, (c) Schluckreflex und pharyngeale Phase, (d) Ösophageale Phase	74
6.17	Modellansicht in der Cycles-Renderengine von Blender mit 20 Samples	74
7.1	Sagittalschnitt des Mundraums bei Bildung des Vokals [i]. (a) nach [Sch10, S. 138], (b) nach [Fiu10, S. 374]	78
7.2	Die Lippen sind im finalen Protoypen prägnanter modelliert, sodass Phone wie [p] (a) und [f] (b) deutlicher dargestellt werden können . . .	79

- 7.3 Myofunktionelle Übungen, die sich mit Schiebereglern kaum realisieren lassen. (a) Kreisen der Zunge an den Lippen, (b) Legen der Zunge zwischen die oberen Frontzähne und Oberlippe, (c) Die Lippen und Mundhöhle werden nicht auf Kollision getestet, (d) Die Zunge lässt sich an die obere Lippe anlegen, da sie mit den oberen Zähnen bereits kollidiert. 80

Kapitel 1

Einleitung

Über die Vor- und Nachteile des Einsatzes digitaler Medien in der Sprachtherapie beschäftigt wurde bereits vor fünfzehn Jahren diskutiert [End99]. Zu diesem Zeitpunkt überwiegen unter den angebotenen Anwendungen diejenigen, die sich an das Konzept des Edutainments hielten. Viele dieser Anwendungen sind inzwischen für das Smartphone oder Tablet als Lernprogramme konzipiert, um dem Patienten das weitere Üben zu Hause zu ermöglichen, oder sie sind Teil der eigentlichen Therapie-Einheit [Sta13]. Eine große Zielgruppe dieser Art von Anwendung sind Patienten mit Sprachstörungen, die sowohl im Sprechen, als auch im Lesen, Verstehen und Schreiben beeinträchtigt sein können. Die meisten dieser Anwendungen präsentieren sich daher in Form von spielerisch gestalteten Übungsaufgaben, wie das Zuordnen von Wörtern zu Bildern.

Das grundlegende Konzept der Behandlung von Patienten mit einer phonetischen Störung, d. h. mit der Beeinträchtigung Laute motorisch zu realisieren, basiert im Gegensatz dazu vor allem auf Artikulationsübungen. Die Therapie setzt sich deshalb zum Ziel, die korrekte Lautbildung und die auditive Wahrnehmung von Fremd- und Eigensprache zu fördern. Dazu greift der Sprachtherapeut üblicherweise auf zweidimensionale Bildmaterialien zurück oder führt den Laut langsam mithilfe seines geöffneten Mundes vor. Taktil-kinästhetische Mittel, durch die die Berührungs- und eigene Körperwahrnehmung des Patienten angeregt wird, oder auditive Ausgaben können zusätzlich dabei helfen, die Sprechbewegung zu veranschaulichen. Zuguterletzt kann der Patient mittels eines Spiegels seine Artikulation auch selbst überprüfen.

Diese Methoden haben jedoch das gemeinsame Problem, dass die eigentliche Stellung der Zunge bei der Bildung eines Lautes nur annähernd verdeutlicht werden kann, da sie schlicht nicht einsehbar ist. Ebenso besteht diese Schwierigkeit in der funktionellen Schlucktherapie, bei der u. a. die Bewegungskontrolle des Schluckreflexes gestärkt werden muss. Trotz der genannten Hilfsmaterialien erscheint es schwierig, einen Laut auf Antrieb für den Patienten verständlich vorzuführen. An die Schwierigkeit kann herangetreten werden, indem eine logopädenseitige Anwendung geschaffen wird, die die verschiedenen Teilprobleme überwindet und so die Visualisierung der Übungen unterstützt.

Die vorliegende Arbeit stellt deshalb prototypisch ein dreidimensionales ComputermodeLL des Vokaltrakts vor, welches sich interaktiv durch den Logopäden bedienen lässt und anhand dessen die Sprechbewegungen und der Schluckvorgang simuliert und visualisiert werden können. Es wird gezeigt, inwieweit die Einbettung dieser interaktiven Visualisierung in die Therapie von Sprech- und Schluckstörungen gelingen kann.

1.1 Motivation

Abhängig von der Einrichtung, in der ein Logopäde arbeitet, ist er im therapeutischen Alltag mit teils stark unterschiedlichen Störungsbildern konfrontiert. So werden beispielsweise Patienten mit Schlaganfall oder neurologischer Erkrankung in der Akutphase häufig stationär in einer Klinik oder einem Rehabilitationszentrum logopädisch behandelt; auch Patienten mit cochleären Implantaten (Innenohrprothesen) werden dort logopädisch betreut. In der ambulanten Therapie treten im Allgemeinen alle Formen von Sprach-, Sprech-, Stimm-, Hör-, und Schluckstörungen auf.

Es liegt auf der Hand, bei heutigen technologischen Möglichkeiten, in der Therapie auf computergestützte Hilfsmittel zurückzugreifen. Eine computerbasierte Therapie ermöglicht neue pädagogische und auch psychologische Herangehensweisen, da eine neue Lernumgebung geschaffen wird und Informationen anders präsentiert werden können. Zusätzlich bereichert sie die klinische Kompetenzen des Therapeuten, wenn ihm neue Evaluationsmöglichkeiten und konstantes Feedback ermöglicht werden. Das Tool kann im Idealfall sowohl bei der Diagnose helfen, als auch audio-visuelle Rückmeldung während der Behandlung geben und den therapeutischen Fortschritt überwachen oder auswerten.

Mithilfe von Fragebögen und Fachgesprächen wurden zunächst Eindrücke vom logopädischen Therapie-Alltag eingeholt. Ziel der Vorab-Umfrage war es zu ermitteln, ob und – falls ja – welche und wie viele digitalen Medien bereits in Therapie-Einheiten verwendet werden. Daraus können erste Annahmen getroffen werden über die Resonanz von solchen Medien in der Sprachtherapie und darüber, ob Zusammenhänge zu dem geplanten Konzept dieser Arbeit bestehen, bzw. welche Anforderungen an eine solche Anwendung gestellt werden.

Von den zwölf befragten Logopäden sind sechs auf einer Schlaganfall-Station und einer neurologischen Rehabilitationsklinik tätig. Der Tätigkeitsbereich der anderen sechs anderen erstreckt sich auf zwei unterschiedliche Praxen, in denen neben der logopädischen Leistung auch Physio- oder Ergotherapie angeboten werden. Die Störungsbilder, die von den Befragten am häufigsten behandelt werden, lassen sich Tabelle 1.1 entnehmen.

Art der Einrichtung	Anzahl befragter Logopäden	Überwiegende Störungsbilder
Klinik (Klinik 1, Klinik 2)	6 (5, 1)	Aphasie, Dysphagie, Dysarthrie, Sprechapraxie, Schlaganfall
Praxis (Praxis 1, Praxis 2)	6 (2, 4)	Aphasie, Dyslalie, Sprechapraxie, Artikulationsstörungen, spezifische Sprachentwicklungsstörungen

Tabelle 1.1: Ausgangsdaten der Umfrage

Auf die Frage, welche Computer-Anwendungen ihnen in der Sprachtherapie bekannt sind, konnten drei Logopäden keine Angaben machen. Zwei Befragte nannten die Anwendung *SpeechTrainer* als Visualisierung von Sprechbewegungen, zwei andere Therapeuten nannten Lernspiele. Alle anderen aufgeführten Anwendungen sind dem Bereich der Diagnostik-Software zuzuordnen, fokussieren sich auf auditive Wahrnehmung bzw. Stimmanalyse oder stellen Untersuchungs-Instrumente mit Monitoring oder Biofeedback dar.

Dennoch geben acht Logopäden auch an, dass bisher keine digitalen Anwendungen in ihrer Einrichtung eingesetzt werden. Nur eine Therapeutin einer Praxis nutzt in ihrer Therapie oben genannte Diagnostik- oder Untersuchungssoftware, drei der vier Einrichtungen verwenden demnach keinerlei computerunterstützten Methoden. Als Gründe für die geringe Akzeptanz von Software-Produkten werden hohe Kosten bildgebender Methoden genannt und eine begrenzte Anzahl an bzw. fehlende Rechner in den Therapieräumen.

Alle befragten Sprachtherapeuten würden trotzdem eine Computer-Anwendung als unterstützend betrachten, da es bei der immer mehr fortschreitenden Technik vor allem zeitgemäß erscheint. Dieses Argument wird gestützt durch die im Schnitt von allen als grundlegend eingeschätzten, allgemeinen Computer-Kenntnisse. Der Patient könne dann je nach Anwendung selbstständig Übungen durchführen und sei nicht auf ein Vorbild angewiesen. Für jüngere Patienten schaffe man durch den spielerischen Umgang mit der Technik mehr Aufmerksamkeit und Motivation für Übungen, besonders, wenn dabei ansprechende Animationen und Belohnungssysteme verwendet würden. Bei der Verwendung von digitalen Medien sei jedoch auch zu berücksichtigen, dass der Einsatz der Anwendung von dem Grad der Einschränkung des Patienten abhängt; für Schlaganfall-Patienten könne sie schnell zu einer Reizüberflutung werden, zumal bei

diesen oft die nötige visuelle oder auditive Wahrnehmung fehle. Gerade ältere Patienten seien zudem im Umgang mit dem Computer wenig geschult und würden „handfeste“ Materialien bevorzugen.

Es wird insgesamt ersichtlich, dass bisher Programme, die dazu dienen konkrete Sprechbewegungen für Patienten verständlich und sichtbar zu vermitteln, überwiegend gar nicht bekannt sind. Faktisch müssen dennoch alle Befragten im Schnitt regelmäßig bis häufig den Bewegungsablauf der Lautbildung in ihrer Therapie einem Patienten mit in der Tabelle genanntem Störungsbild vorführen.

1.2 Zielsetzung

Ziel der vorliegenden Arbeit ist es, Logopäden ein therapeutisches Werkzeug zu bieten, das sie bei der Visualisierung einer beliebigen Lautbewegung und des physiologischen Schluckakts unterstützt. Auf Grundlage eines anatomisch korrekten, dreidimensionalen Modells der oro-pharyngo-laryngealen Region sollen die Bewegungsvorgänge realistisch simuliert werden. Dem Therapeut soll die Möglichkeit geboten werden, mit diesem Modell zu interagieren, d. h. vor allem spezifische Laute und den Schluckvorgang automatisch vorführen zu können. Der Patient erhält durch den Vergleich seiner Lautbildung mit der des Modells indirekt eine Selbstreflexion in Echtzeit. Mithilfe dieser prototypischen Anwendung soll der Aspekt untersucht werden, ob und inwieweit eine solche Anwendung langfristig einen Mehrwert zu den üblichen Hilfsmitteln der Sprachtherapie darstellen kann und welche Störungsbilder davon profitieren könnten. Dafür werden iterativ die Anforderungen von Logopäden an eine solche Applikation berücksichtigt. Teile der Arbeit sollen sich dabei auch einem Vergleich mit der *SpeechTrainer*-Software der Universität Aachen oder computeranimierten Tutoren wie *Baldi* und *ARTUR* unterziehen.

1.3 Aufbau der Arbeit

Der Aufbau dieser Masterarbeit gestaltet sich wie folgt: Das nächste Kapitel fasst die theoretischen Grundlagen aus Anatomie und Linguistik zusammen. Im folgenden Kapitel werden darauf aufbauend die für diese Arbeit relevanten logopädischen Störungsbilder vorgestellt und auf gängige Therapiemethoden für die Störungen eingegangen. Sie stecken das für diese Arbeit potentielle Klientel ab. Im darauffolgenden Kapitel 4 werden verwandte Forschungsansätze dargelegt, um anschließend in Kapitel 5 das Konzept dieser Arbeit vergleichend vorzustellen. Kapitel 6 beschreibt den ersten Prototypen und zieht mithilfe einer Anforderungsanalyse ein erstes Fazit. Darauf aufbauend wird der finale Prototyp präsentiert. Dieser wird in Kapitel 7 als Experteninterview abschließend

evaluiert. Es folgt das Resümee und der Ausblick, der sich den weitergehenden Aspekten dieser Arbeit widmet.

Um den Textfluss nicht zu stören, werden bei Patienten und Berufsbezeichnungen die grammatikalisch maskuline Form gewählt. In diesen Fällen sind jedoch immer Frauen und Männer gemeint.

Kapitel 2

Grundlagen der Anatomie und Lautlehre

Im Kontext der Lautlehre wird sowohl ein grundlegendes anatomisches und physiologisches Wissen über die biomechanischen und motorischen Vorgänge des Sprechakts vorausgesetzt, als auch Wissen über das Sprachsystem im linguistischen Sinn. Phonetik und Phonemik sind so in einem direkten Zusammenhang zu betrachten.

Der nächste Absatz geht zunächst auf die physiologischen Grundlagen der Lautbildung und des Schluckens ein. Darauf aufbauend wird die Phonetik aus artikulatorischer Sicht betrachtet.

2.1 Physiologische Sicht

Die Bildung von Sprechlauten setzt Artikulationsorgane voraus, die vom Nervensystem gesteuert und koordiniert werden. Als Artikulationsorgane bezeichnet man diejenigen Organe, die bei der Artikulation und Lautbildung zum Einsatz kommen.

Die Öffnung des Kiefers wird u. a. durch die Kaumuskeln sowie die oberen und unteren Zungenbeinmuskeln herbeigeführt. Dabei spricht man bei der gesamten horizontal stehenden Muskelplatte zwischen Zungenbein und Unterkiefer vom muskulären Mundboden, der sowohl die oberen Zungenbeinmuskeln beinhaltet als auch Teile der Zungenmuskulatur. Verdeutlicht sind die zum Zungenbein gehörigen Muskelgruppen in Abb. 2.1a und Abb. 2.1b.

Die Zunge spielt neben der Lautbildung eine wichtige Rolle bei der Nahrungsaufnahme, als Sinnesorgan und für den Saugakt, bei dem sie den nötigen Unterdruck erzeugen kann. Sie ist am Mundboden befestigt und besteht zu einem großen Teil aus quergestreifter Skelettmuskulatur (Abb. 2.2). Sie ist das einzige quergestreifte Skelettmuskel-

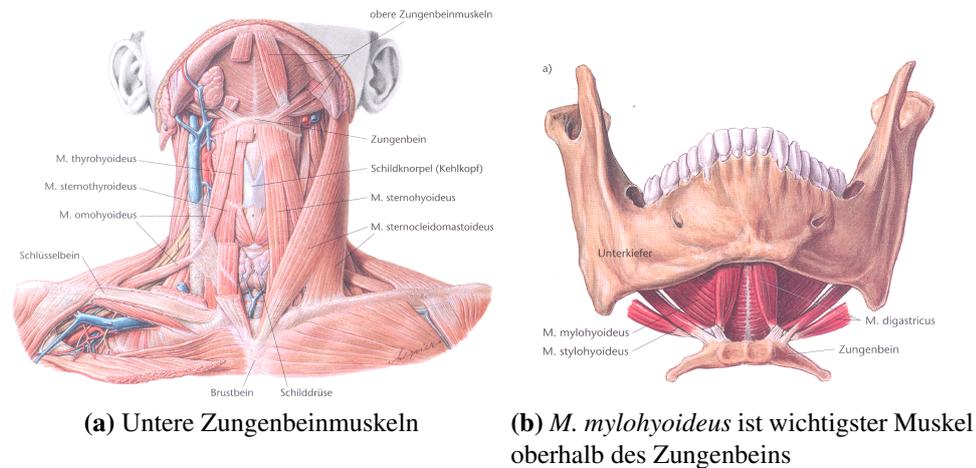


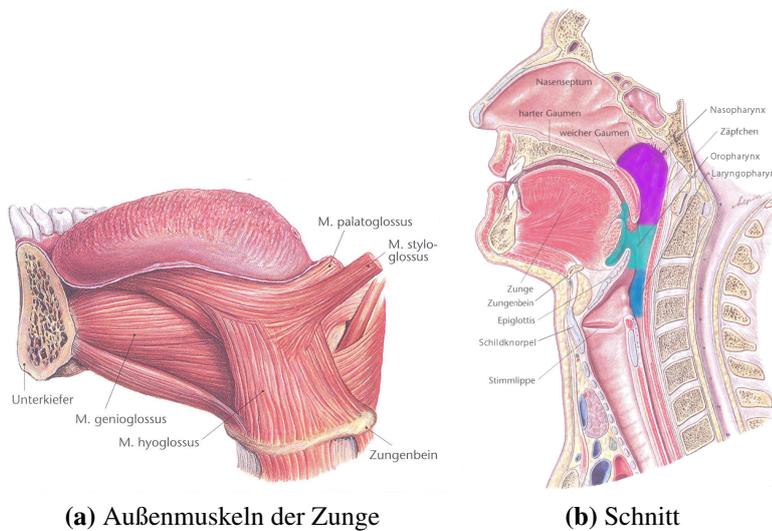
Abbildung 2.1: Ansicht von vorne auf die unteren und oberen Zungenbeinmuskeln nach Schindelmeiser [Sch10, S. 85, 86]

system, das sich bei einer Kontraktion nicht nur verkürzen, sondern auch verlängern kann (Herausstrecken der Zunge). Die Muskulatur der Zunge lässt sich in Binnen- und Außenmuskeln gliedern und dienen der Verformbarkeit und Beweglichkeit der Zunge. Die Binnenmuskeln verlaufen innerhalb der Zunge und lassen die Form der Zunge in alle drei Raumrichtungen verändern. Die Binnenmuskeln, die in Längsrichtung verlaufen, verkürzen die Zunge in Längsrichtung, diejenigen in Querrichtung verkürzen sie in Querrichtung und können die Zunge dadurch schmaler machen. Die vertikalen Muskeln machen die Zunge flacher.

Die Außenmuskeln führen von benachbarten Skeletteilen bis zur Zunge. Der Muskel *M. genioglossus* (Kinn-Zungen-Muskel) zieht die Zunge nach vorne, *M. hyoglossus* (Zungenbein-Zungen-Muskel) nach hinten unten und *M. styloglossus* (Griffelfortsatz-Zungen-Muskel) nach hinten oben. *M. palatoglossus* (Gaumen-Zungen-Muskel) erstreckt sich bis zur Zungenwurzel und spielt vor allem beim Schluckakt eine Rolle (Abb. 2.2a).

Die Zunge selbst besteht aus anatomischer Sicht aus einer Zungenspitze, einem Zungenkörper und einer Zungenwurzel, welcher bei einem geöffneten Mund nicht sichtbar ist. Der Zungenrücken ist die bei normaler Zungenlage gesamte Zungenoberfläche und beinhaltet die Zungenspitze und die seitlichen Zungenränder bis zur V-förmigen Grenzfurche, die die Grenze zwischen den vorderen zwei Dritteln und dem hinteren Drittel der Zunge bildet [Sch10].

Der Gaumen besteht aus zwei Dritteln hartem und ein Drittel weichen Gaumen, welcher das Gaumensegel (Velum) mit Zäpfchen (Uvula) beinhaltet. Bei noch unzerkauter Nahrung ist das Velum gesenkt und der Racheneingang gesenkt. So kann die Nah-



(a) Außenmuskeln der Zunge

(b) Schnitt

Abbildung 2.2: Seitliche Ansicht auf die äußeren Muskeln der Zunge nach Schindelmeiser [Sch10, S. 132, 136] und den oro-pharyngo-laryngealen Bereich nach Bartolome [BSM14, S. 13]

ung nicht vorzeitig Abrutschen. Während des Schluckvorgangs trennt dann das Velum den Mundraum vom Nasenraum, damit dort keine Nahrung hingelangen kann. Dies geschieht auch bei der Artikulation, außer bei der Bildung von nasalen Konsonanten, da hier die Luft durch die Nase entweichen muss.

Im Rachen (Pharynx) verbinden sich Luft- und Speiseröhre. Durch Kontraktion seiner Muskeln lässt sich die Nahrung in Richtung Schlund bewegen. Er erstreckt sich daher bis zum Speiseröhreneingang und lässt sich in drei Bereiche teilen (Abb. 2.2b): Der Nasopharynx befindet sich über dem Gaumensegel vor der Nasenhöhle und geht bis zum Rachendach (lila). Der Oropharynx befindet sich darunter und befindet sich auf Höhe der Mundhöhle, zu der er sich öffnet (hellblau). Der Laryngopharynx geht vom Kehledeckel (Epiglottis) bis zum Ösophagus-Beginn und Kehlkopf (blau). Als Teil des Ansatzrohrs, das alle Hohlräume des Kehlkopfs oberhalb der Stimmritzen umfasst (oro-pharyngo-laryngealer Raum), ist der Rachen auch an der Lautbildung beteiligt.

Die unteren Atemwege werden durch den Kehlkopf (Larynx) geschützt und wird begrenzt durch die Epiglottis. Weiterer Teile des Kehlkopfskeletts sind auch der beim Schlucken beteiligte Ringknorpel, die beiden Stellknorpel und Schildknorpel. Sie beteiligen sich am Verschließen des Atemwegs während des Schluckens. Die zum Kehlkopf gehörigen Stimmritzen dienen nicht nur der Lautbildung, sondern werden beim

Schlucken zusätzlich geschlossen [BSM14, S. 12ff.].

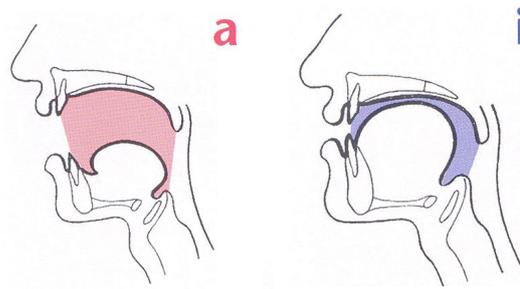
2.1.1 Schluckvorgang

Die Komplexität des Schluckens von aufgenommener Nahrung lässt sich üblicherweise auf vier Phasen reduzieren. Bei deren Beschreibung soll sich hier insbesondere auf das Verhalten der Zunge und des Gaumensegels konzentriert werden [BSM14, S. 24].

- In der oralen Vorbereitungsphase werden über Kaubewegungen feste Speisen zerkleinert und mit Speichel vermischt (genannt Bolus), auf das vordere bis mittlere Zungendrittel gebracht und von der Zunge umschlossen. Die Zunge bildet eine Schüssel. Der hintere Zungenteil ist umschlossen. Die Zungenspitze und -ränder liegen alveolar an. Das Gaumensegel ist nach vorne gesenkt und schließt den Mundraum ab, damit keine Speisematerialien schon vor dem Schluckreflex in den Oropharynx gelangen.
- Während der oralen Phase hebt sich das Velum und die Zungemitte bildet eine Furche. Durch Kontraktion von extrinsischen und intrinsischen Zungenmuskeln wird der Bolus weiter nach hinten transportiert. Dieser kann anschließend über die gesenkte Hinterzunge rampenartig in den Oropharynx gleiten.
- Der eigentliche Schluckreflex setzt in der pharyngealen Schluckphase ein. Das Gaumensegel schließt nun mit Vorwölbung der Rachenwand den Nasenraum komplett ab, damit die Nahrung nicht dorthin gelangt (velopharyngealer Verschluss). Wenn der Bolus den Zungengrund im Hypopharynx erreicht hat, verschiebt sich die Zungenbasis kurzzeitig in Richtung Rachenwand. Gleichzeitig bewegen sich Zungenbein und Kehlkopf nach oben und vorne, womit der Rachenraum für den Bolus geweitet wird. Für die Kehlkopfhebung nähert sich der Aryknorpel an die Epiglottis an, welche horizontal durchkippt und den Kehlkopfeingang unvollständig schließt. Dadurch können Speiseteile nicht in die Luftröhre gelangen. Zusätzlich sind auch die Stimmlippen geschlossen.
- Schließlich gelangt in der ösophagealen Phase der Bolus nach Öffnung des sog. Ösophagusphinkters durch die Speisröhre in den Magen.

2.1.2 Artikulation

Im Gegensatz zur Stimmbildung, bei der der Kehlkopf von größter Bedeutung ist und mit ihm Tonhöhe, Frequenzumfang, Stimmklang und Lautstärke beeinflusst werden können, wird die eigentliche Lautbildung (Artikulation) mithilfe der Artikulationsorgane im Ansatzrohr gesteuert [Sch10, S. 107].



(a) Resonanzraum bei [a] (b) Resonanzraum bei [i]

Abbildung 2.3: Resonanzraum (rot und blau) bei Bildung des Vokals [a] und [i] nach Schindlmeiser [Sch10, S. 138]

Prinzipiell spielen vom Ansatzrohr die Lippen, der Unterkiefer und das Gaumensegel eine große Rolle bei der Lautbildung. Für die exakte Bildung eines Lautes sind dabei Lippenöffnung und die Formung der Lippen sowie der Zahnreihenabstand (Stellung des Unterkiefers) von enormer Wichtigkeit, da sie sonst resonatorische und artikulatorische Fehler nach sich ziehen [Fiu10, S. 62ff.]. Das Velum (Gaumensegel) kann die Nasenhöhle vom Mund-Rachen-Raum abgrenzen, indem es sich an die Rachenrückwand legt, oder es kann zusammen mit dem Zungenrücken die Mundhöhle abschließen. Beim Schlucken wird damit verhindert, dass die Nahrung durch die oberen Luftwege gelangt. Durch das Heben und Senken wird so der Phonationsstrom bei der Artikulation gelenkt.

Hauptakteur bei der Lautbildung ist jedoch die Zunge, da sie sich von allen Artikulationsorganen am meisten bewegen und formen lässt. Aus phonetischer Betrachtung lässt sich die Zunge in einen Zungenkranz (Zungenspitze mit Seitenrändern), ein Zungenblatt (vorderes Drittel vom Zungenkörper ohne Zungenspitze) und Zungenrücken (mittleres Drittel des Zungenkörpers) einteilen. Dadurch wird eine differenzierte Beschreibung der Zungenlaute möglich. Wichtig in der deutschen Sprache ist der flexible Kontakt von Zungenspitze mit den palatinalen Flächen und lingualen Flächen der Frontzähne. Die Mundhöhle dient als Resonanzraum, der unterschiedlich geformt wird. Beispiele für den Mund- und Rachenraums bei Bildung des Phons [i] und [a] zeigen Abb. 2.3a und 2.3b.

Einer Störung der Lautbildung können aus anatomischer Sicht verschiedenste Ursachen zu Grunde liegen, wobei es sich meist nicht nur um eine Störung handelt. Diese reichen von einer Muskellähmung des Mundringmuskels, die auch die Nahrungs- und Flüssigkeitsaufnahme behindert, über ein Ausfallen des 12. Hirnnervens, der die Zungenmuskeln erregt, bis hin zu Fehlbildungen und Entwicklungsstörungen (Lippen-, Kiefer-, Gaumenspalten) [Sch10, S. 121, S. 135].

2.2 Einteilung von Sprachlauten

Ein Phon in der Linguistik beschreibt den kleinsten unterscheidbaren Einzellaut und wird in eckigen Klammern als Zeichen des Internationalen Phonetischen Alphabets (IPA) notiert. Es ist zu beachten, dass ein Phon während des Sprechens nie einzeln auftritt, sondern immer in Verbindung mit anderen Einzellauten stehen, die sich in der Artikulationsbewegung gegenseitig beeinflussen (Koartikulation). Eine Differenzierung der Laute im Deutschen besteht aus Vokalen und Konsonanten.

Vokale lassen sich anhand des Rundungsgrads der Lippen, Öffnungsgrads des Kiefers und der Zungenlage einteilen. Ein vereinfachtes Vokalviereck (ohne [ɐ], [i:], [e:], [o:], [u:], [y:], [ø] und Diphthonge) verdeutlicht die Bildung der Vokale durch Zungenhöhe und Zungenlage in Mundraum in Abbildung 2.4a. Bei Vokalpaaren wie i • y steht das linke Phon für den ungerundeten Vokal, das rechte für den mit gerundeten Lippen.

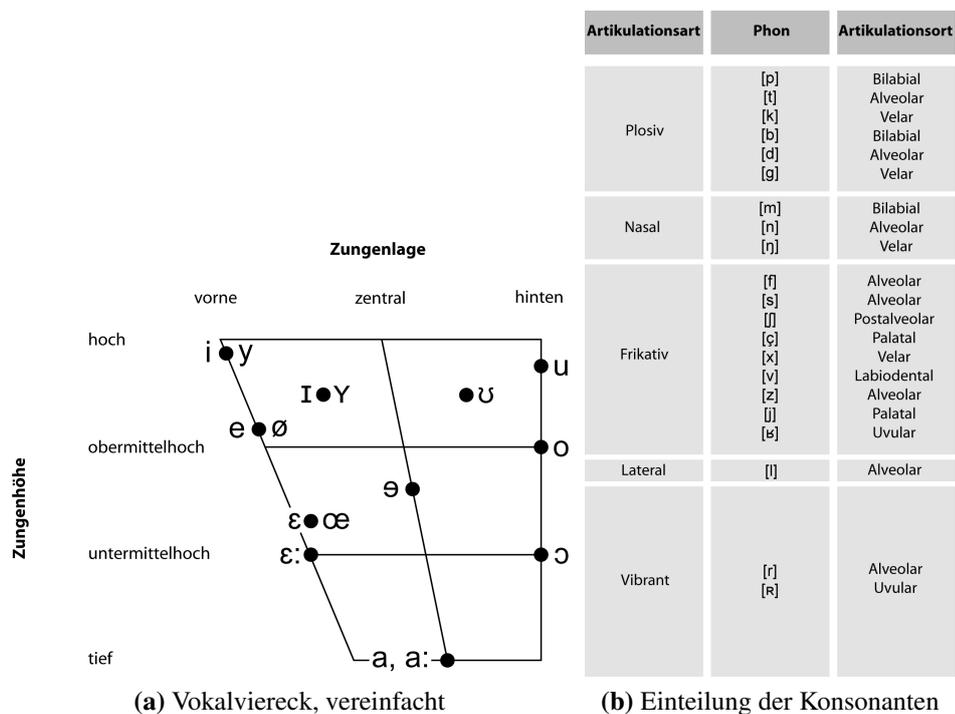


Abbildung 2.4: Die Unterteilung von Vokalen und Konsonanten im Deutschen über Zungenhöhe und Zungenhebung bzw. Artikulationsort und Artikulationsart nach [WZ11, S. 4, 5]

Die Unterteilung von Konsonanten geschieht neben dem Artikulationsorgan meist über den Artikulationsort oder den Artikulationsmodus. Als Artikulationsort wird diejenige

Stelle bezeichnet, deren Muskelspannung bei Bildung des Lauts verändert wird (z. B. Zäpfchen) oder wo der Laut gebildet wird (z. B. Lippe oder Zunge). So bezeichnet man beispielsweise Laute als Bilabiale, wenn sie an fest geschlossenen Lippen gebildet werden oder als Labiodentale, wenn sie an Lippen und Zähnen gebildet werden. Im Gegensatz dazu beschreibt man durch den Artikulationsmodus die Art und Weise, durch die der Sprechlaut gebildet wird [Sch10, S. 138-140]. Ein Plosiv stellt beispielsweise einen Explosivlaut dar, bei dem der Mund vollständig geschlossen ist, und der Phonationsstrom angestaut wird, bevor er abrupt (sprengend) freigegeben wird. Die Artikulationsart und -stelle einiger Konsonanten finden sich in Tabelle 2.4b.

Bei der Bildung von Konsonanten sind Zunge, Lippen, Wangen, weicher Gaumen und Stimmritze beteiligt. Hierbei wird der Luftweg eng gestellt, woran die Ausatemluft nicht wie bei den Vokalen unbehindert herausströmen kann, sondern vorbei geleitet wird. Anschließend werden verschlossene Bereiche des Luftweges „aufgesprengt“. Konsonanten können stimmlos (z. B. [ch] und [f]) oder stimmhaft sein (z. B. [l] und [n]). Erstere entstehen als reine Geräusche an den Artikulationsorten, letztere sind Klanggemische, bei denen oft auch der vom Kehlkopf erzeugte Stimmklang hinzukommt. Die Bezeichnung für einen mit großer Intensität gesprochenen Konsonant ist eine Fortis, der sanft gesprochene eine Lenis ([p]/[b]).

Beispiele für die Aussprache eines Phons als Vokal oder Konsonant finden sich in Anhang A.2.

Kapitel 3

Grundlagen der Sprachheilkunde

Im Bereich des Sprachheilwesens wird in Deutschland eine besondere Konstellation ersichtlich, bestehend aus unterschiedlichen sprachtherapeutischen Berufsgruppen mit jeweils eigenen Abschlüssen und Schwerpunkten. Im Gegensatz zu den europäischen Nachbarländern dürfen hier auch nichtakademische Berufsgruppen im Heilwesen tätig sein, wie dies der Fall bei Logopäden oder Atem-, Sprech- und Stimmlehrern ist, die an Fachschulen staatlich geprüft werden [SRH13, S. 22–25]. Aufgrund des Strukturwandels der letzten zwei Jahrzehnte und der daran gekoppelten Vielzahl an Überschneidungen der einzelnen Berufszweige, betreffen nachfolgende Erläuterungen nur den Beruf der akademischen Sprachtherapie.

Im Kontext der sprachtherapeutischen Berufe ist die akademische Sprachtherapie eine eigenständige klinische, und die jüngste Fachdisziplin. Sie entstand aus der Sprachheilpädagogik in Verbindung mit der Klinischen Linguistik, Patholinguistik und Klinischen Sprechwissenschaft, was ihren Bezug nicht nur zur Medizin sondern im Besonderen auch zur Linguistik offenbart [Gro14]. Im Fokus ihrer Beschäftigung liegen Prävention, Diagnostik, Behandlung und Erforschung von Sprach-, Sprech-, Stimm-, Hör-, Schluck- und Kommunikationsstörungen. Die Sprachtherapie wird dabei als Integrationswissenschaft aufgefasst (Abb. 3.1), da sie als interdisziplinäres Fach auch Bezug nehmen muss auf fachübergreifendes Wissen wie beispielsweise aus der Medizin oder der Psychologie [Gro12].

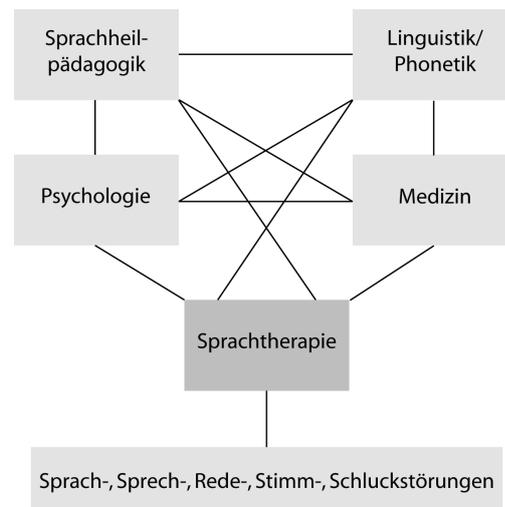


Abbildung 3.1: Aufgabengebiete der Sprachtherapie nach Grohnfeldt [Gro14, S. 17]

3.1 Störungsbilder

Das Klientel der Sprachtherapie umfasst Menschen aller Altersgruppen mit Sprach-, Sprech-, Rede-, Stimm- und Schluckstörungen unterschiedlichen Ausmaßes. Diese vereinfachte Klassifikation in fünf Komponenten hilft dabei, die auftretenden Symptome genauer einzuordnen. Sie stellt mehr ein grundlegendes Hilfsmittel für ein individuell abgestimmtes Therapievorgehen dar, denn eine statische Untergliederung.

Demnach existiert eine Reihe potentieller Erscheinungsformen von Störungen, die nicht eindeutig voneinander abzugrenzen sind und zu Mischformen führen. Zusätzlich lässt sich „normales“ und auffälliges Sprachverhalten nicht immer klar differenzieren. Die Frage, wie häufig welche Störungen in der Gesellschaft auftreten, ist deshalb schwer zu beurteilen und unterliegt obendrein dem gesellschaftlichen Wandel [Gro12, S. 73ff.].

Im Fokus dieser Arbeit liegen insbesondere die Klasse der Sprach- und Sprechstörungen mit Auswirkungen auf die Sprechmotorik und -koordination, und die der Schluckstörungen. Dies steht im Einklang mit der Zielsetzung der Arbeit, die Bewegungsabläufe der Zunge beim Sprechen und Schlucken zu simulieren.

Die Unterschiede der einzelnen logopädischen Störungsbilder werden im kommenden Abschnitt kurz verdeutlicht und in Tabelle 3.1 abschließend festgehalten. Aus Gründen der Übersicht wird dabei nur auf die wichtigsten Störungsbilder eingegangen.

3.1.1 Störungen der Sprach- bzw. Lautsprachentwicklung

Diese Entwicklungsstörungen sind der Oberbegriff für Störungen während des frühkindlichen Spracherwerbs und offenbaren sich am häufigsten erst bei Kindern im Vorschul-

und Grundschulalter. Im Allgemeinen lassen sich die Ebenen einer Sprache durch folgende linguistische Termini beschreiben:

- Phonetik
- Phonologie
- Syntax und Morphologie
- Semantik und Lexik.

Störungen in den letzten beiden Ebenen stellen eine Beeinträchtigung bei der Erlernung von grammatikalischen Regelsystemen und Wortschätzen dar und sollen hier nicht weiter betrachtet werden. Gerade die phonetischen Störungen sind als Artikulationsstörungen jedoch von höherem Interesse für diese Arbeit; sie repräsentieren in dem Klassifizierungsmodell nach Dodd¹ zusammen mit den phonologischen Störungen die Gruppe der **Aussprachestörungen** [Gro12, S. 85].

Sie bedeuten zunächst Abweichungen im Lauterwerb in dem Umfang, dass die Betroffenen im Allgemeinen in ihrer Sprachentwicklung zurückliegen. Die Ursachen dafür können u. a. eine unzureichende Koordination oder eine Bewegungsstörung der Muskulatur [Gro12, S. 87] oder organischer Natur sein, z. B. in Form von neurologisch bedingten Dysarthrien oder Dyspraxien (s. Abschnitt 3.1.2). Sie können auch funktioneller Natur sein. In diesem Fall ist keine konkrete organische Ursache bekannt, dafür beispielsweise eine genetische oder psychische [Gro14, S. 176]. Da es insgesamt aber eine Reihe unterschiedlicher Abweichungen in der Aussprache geben kann, die eine nähere Unterteilung erfordern, bleibt die Klasse der Aussprachestörungen lediglich ein weiterer Oberbegriff.

Eine **phonetische Störung** ist eine Sprechstörung und äußert sich beim Betroffenen so, dass es ihm nicht möglich ist, einzelne Phone korrekt umzusetzen. Ein Phon ist das kleinste Lautsegment, das unabhängig von seiner Funktion oder Bedeutung innerhalb eines Sprachgefüges isoliert gebildet werden kann. Verbreitetes Beispiel für eine Sprechstörung ist der als „Lispeln“ bekannte Sigmatismus, bei dem die Laute [s], [z], [ʃ] und [ç] fehlerhaft gebildet werden. Anderes Beispiel ist der Kappazismus, bei dem der Laut [k] nicht korrekt gebildet oder durch den Laut [t] ersetzt wird,

Die phonetische Störung kann einzeln auftreten, steht aber nicht selten im Zusammenhang mit einer **phonologischen Störung**. Bei dieser liegt das Gegenteil vor: Phoneme können im Sprechfluss nicht an korrekter Stelle verwendet, also im Gefüge nicht wie vorgesehen verbunden werden. Die Funktion dieses Lauts ist folglich nicht bekannt.

¹s. Dodd, B.: *Differential diagnosis and treatment of children with speech disorder*, 1995

3.1.2 Zentrale Sprach- und Sprechstörungen

Verbale Entwicklungsdyspraxien (VED) sind Störungen, die die Sprechbewegungs-Planung und -Programmierung in der Phase des Erstspracherwerbs betreffen [Gro14, S. 183]. Die Beeinträchtigung besteht darin, die für die Artikulation notwendigen Bewegungen nicht bewusst und kontrolliert ausführen zu können, weder räumlich noch zeitlich. Daraus resultierend wird sowohl der Lauterwerb als auch das Erlernen der Grammatik erschwert.

Die Auffälligkeiten bei einer VED variieren je nach Kind unterschiedlich stark. Zum einen sind bei einigen Kindern keine Lall- und Plapperphase in der Säuglingszeit gegeben, sodass der Sprechbeginn dementsprechend verspätet ist. Kommunikationsversuche finden dann nur mithilfe von Gestik, Mimik und Zeigen statt. Manche Kinder können nur Vokale produzieren. Ein Kernsymptom ist es auch, wenn die Kindern schon beherrschte Laute und Wörter nicht mehr bilden können. Die Äußerungen sind kaum verständlich. Als typischer Indikator für eine VED sind jedoch inkonsistente Lautfehlbildungen, bei denen Laute vertauscht und kompensatorisch verwendet werden [Gro14, S. 184f.].

Erworbene, neurologisch bedingte Störungen, die ebenso die Ausführung von Sprechbewegungen hemmen, sind zum einen Sprechapraxien und davon abgegrenzt Dysarthrien. Die Symptomatik der **Dysarthrie** weist häufig unterschiedlich starke Funktionsstörungen in der Sprachatmung, Stimme und Artikulation auf, wenn eine beeinträchtigte Atem- und Kehlkopfmuskulatur vorliegt und/oder motorische Anteile des Nervensystems geschädigt sind. Ursache können akute Hirnschädigungen wie nach einem Schlaganfall sein, neurodegenerative Erkrankungen wie die Parkinson-Krankheit oder Entzündungen des Nervensystems wie Multiple Sklerose. Sie gehört zur am häufigsten neurologisch verursachten Kommunikationsstörung [Gro14, S. 275f.].

Bei einer **Sprechapraxie** sind sprechmotorische Routinen beeinträchtigt. Konkrete Muskelbefehle und ihre zeitlich-räumlichen Bewegungsinformationen werden nach der phonologischen Verarbeitung des Gesprochenen nicht mehr korrekt umgesetzt. Die Störung beeinflusst demnach nur die Sprechbewegungen. Zentrale Symptome sind Störungen der Lautbildung und phonematische sowie phonetische Fehler. Die Fehlermuster äußern sich nicht immer identisch, sondern treten inkonsistent auf. Eine Sprechanstrengung mit ständigen Korrekturversuchen führt zu Anspannungen im Gesichts- und Halsbereich. Ursachen dieser Störung sind hauptsächlich Erkrankungen, die die Blutgefäße des Gehirns betreffen, wie Hirninfarkte oder -Blutungen [Gro14, S. 281ff.].

Eine **Aphasie** ist eine weitere Sprachstörung, bei der es nach Verletzung der linken Gehirnhemisphäre (Sprachregion), die zu einer verschieden starken Beeinträchtigung der Sprachfähigkeit führt. Die Fähigkeit Sprache zu produzieren, zu verstehen, Lesen

oder Schreiben sind begrenzt. Sie ist demnach neurologisch bedingt und tritt häufig nach einem Schlaganfall, aber auch infolge eines Hirntumors auf.

3.1.3 Rhinophonien

Rhinophonie ist auch unter Näseln bekannt und stellt eine Resonanzstörung dar, die den Stimmklang und die Artikulation beeinflussen und mit erhöhter oder verringerter Nasalität zusammenhängen. Fehlbildungen wie LGKS-Spalten² können häufig zu einer kompensatorischen Artikulation führen, bei der ein Ziellaut komplett durch eine ausgleichende Bewegung ersetzt wird, um ein deutlicheres Sprechen zu erreichen [Gro14, S. 286ff.].

3.1.4 Schluckstörungen

Die Schluckstörung (Dysphagie) ist von obigen Störungen des Sprechens/Sprache abzugrenzen, sie beeinträchtigt die lebenswichtige Funktion des Transports von Nahrung, Flüssigkeit und Speichel von der Mundhöhle in die Speiseröhre. Ursachen können dafür sensomotorischer Natur sein, d. h. der Schluckvorgang kann nicht korrekt in seinen Bewegungen gesteuert werden, oder auch psychogener Natur.

3.2 Therapie

Einer Therapie geht zunächst die Ermittlung einer potentiellen Störung voraus (Diagnose). Weitergehend gibt sie aber auch die nötigen Informationen, um anschließende therapeutische Maßnahmen einzuleiten, zu planen und zu gestalten.

Die sprachtherapeutische Diagnostik stellt sich als einen fortlaufenden Prozess dar. Zuerst wird z. B. in Form eines Gutachtens eingeschätzt, ob eine Behandlung nötig ist; darauf aufbauend werden spezifische diagnostische Maßnahmen geplant und eingeleitet. Auch während der Therapie werden kontinuierlich weitere diagnostische Erkenntnisse gewonnen und die Ziele und Methoden an diese angepasst. Vor allem bei Kindern wird davon ausgegangen, dass sich ihre Fähigkeiten mit fortschreitendem Alter vergrößern, was bestmöglich unterstützt werden muss. Deshalb sind nicht nur die sprachlich-kommunikativen Fähigkeiten von Bedeutung, sondern vielmehr die Gesamtentwicklung des Klienten unter Berücksichtigung seiner Umwelt. Auch Aspekte wie Sensorik und Motorik sowie kognitive Faktoren werden beachtet. Für die Planung der anschließenden Therapie sind demnach individuelle Lernvoraussetzungen grundlegend, wofür u. a. auch

²Lippen-Kiefer-Gaumensegel-Spalten

STÖRUNGEN DER SPRACHENTWICKLUNG	<i>Sprache</i>	<i>Sprechen</i>
Aussprachestörung (phonetisch)		x
Aussprachestörung (phonologisch)	x	(x)
Sprach-Entwicklungsstörung (phonologisch)	x	
Semantisch-lexikalische Störung	x	
Sprach-Entwicklungsstörung bei kognitiven Behinderungen	x	
Sprach-Entwicklungsstörung bei Mehrsprachigkeit	x	
Sprach-Verständnisstörungen	x	
Sprach-Entwicklungsstörung bei hochgradiger Schwerhörigkeit und Gehörlosigkeit	x	
ZENTRALE SPRACH- UND SPRECHSTÖRUNGEN		
Aphasien	x	
Sprachabbau bei Demenz	x	
Störungen der Sprechmotorik (Dysarthrien/Dysarthrophonien, Sprechapraxien, verbale Entwicklungsdyspraxien)		x
RHINOPHONIEN		
		x

Tabelle 3.1: Störungsbilder der Komponenten Sprache und Sprechen, Auszug aus Grohnfeldt [Gro12, Tab. 7, S. 74]. Schluckstörungen stellen ein eigenes Störungsbild dar und sind hier nicht aufgeführt

das soziale und familiäre Umfeld des Patienten analysiert und sein Lebenslauf rekonstruiert wird [Gro14, S. 342ff.].

Abhängig vom Störungsbild kommen dann nach der Diagnose angepasste therapeutische Vorgehensweisen zum Einsatz, die auch auf die pädagogische Sprachförderung zurückgreifen. Wie bereits erwähnt, erfolgt die Planung und Durchführung von Therapiesitzungen systematisch, ihr Verlauf wird dabei beobachtet und ggf. revidiert.

Definitionsgemäß sollte der Therapeut in erster Linie solche Maßnahmen treffen, die sowohl die vorliegende Sprachstörung als auch die damit einhergehenden psychosozialen Konsequenzen des Patienten lindern oder langfristig gar beseitigen. Deshalb ist es ein wesentliches Ziel der Sprachtherapie, diejenigen Kompetenzen als Therapeut zu entwickeln, die zur „selbständigen und adäquaten Problembewältigung der Betroffenen führen“ [Gro14, S. 30]. Die subjektive Wahrnehmung und die eigene Auseinandersetzung mit der Störung haben beim betroffenen Klienten oftmals eine hohe Relevanz, sodass die Beeinträchtigung leicht psychosoziale Auswirkungen (Erwartungsängste, ne-

gative Bewertungen durch die Umwelt) hervorrufen. Diese Aspekte sind essentiell für die Sprachtherapie, da sie von der objektiven Beurteilung der Störung durch den Therapeuten abweichen können [Gro12, S. 17].

So sollen einerseits durch standardisierte Testverfahren objektive Erfolge der Maßnahmen gewährleistet werden, andererseits müssen aber auch eine positive subjektive Zufriedenheit beim Klienten angestrebt und seine biographischen Faktoren berücksichtigt werden [Gro12, S. 75ff.]. Die Therapiesitzungen werden nicht nur deshalb interaktiv in Zusammenarbeit zwischen Therapeut und Klient konzipiert.

Die nachfolgenden zwei Abschnitte befassen sich kurz mit gängigen Therapieverfahren, beschränkt auf die bereits vorgestellten Störungsbilder (s. Kapitel 3.1).

3.2.1 Diagnoseansätze

Eine störungsspezifische sprachtherapeutische Behandlung setzt eine differenzierte Diagnostik voraus. Einige Methoden der Diagnostik sollen kurz umrissen werden, da sie einerseits den Anstoß zur Therapieplanung geben und damit die Einbettung dieser Arbeit in den Prozess der Sprachtherapie begründen. Andererseits stellen sie bereits mögliche Übungsmaterialien vor.

So ist eine erste Diagnostik bei Aussprachestörungen, Konsonanten oder Konsonantenverbindungen in einem Wort überprüfen zu lassen oder mit ihnen das phonologische System des Klienten zu ermitteln. Kinder benennen vorgegebene Bilder, die ihrem Wortschatz entsprechen. Es wird geprüft, ob ein gleiches Wort vom Kind immer wieder identisch ausgesprochen werden kann [Gro14, S. 178f.]. Für die funktionsorientierte Diagnostik von Dysarthrien existieren Untersuchungsmaterialien mit unterschiedlichen Aufgaben. Mit phonetisch strukturierten Wortlisten, Sätzen zum Nachsprechen und Lesetexten werden Symptome von Sprechatmung, Stimme, Artikulation, Resonanz und Prosodie geprüft. Bei nichtsprachlichen Bewegungsaufgaben müssen Zunge oder Lippen einzeln bewegt werden. Auditive Auswertungen des Gesprochenen sind ein Standard der klinischen Dysarthriediagnostik. Diagnosemöglichkeiten bei Sprechapraxien können sowohl eine Untersuchung der Spontansprache umfassen als auch systematische Untersuchungsverfahren der Sprechmotorik mit vorher geordnetem und phonetisch kontrolliertem Material. So kann u. a. der Einfluss der Silbenkomplexität auf ein Fehlermuster ermittelt werden. Als Beispiele solcher Verfahren sind „Hierarchische Wortlisten“ zu nennen, die vom Patienten nachgesprochen werden sollen. Phonetische Signale werden durch auditive Verfahren analysiert [Gro14, S. 282f.].

3.2.2 Therapiekonzepte

Die hier aufgeführten Methoden stehen im Zusammenhang mit der Zielsetzung dieser Arbeit und zeigen deshalb Einsatzmöglichkeiten für die Anwendung auf. Üblicherweise gibt es eine Vielzahl an Therapie-Methoden, die ständig erweitert und überarbeitet, und die oftmals nicht nur einzeln angewandt werden.

Aussprachestörungen

Der älteste bekannte Therapieansatz im deutschsprachigen Raum bei Artikulationsstörungen findet sich in der „klassischen Artikulationstherapie“ nach Van Riper, auf den einige weitergehende Therapiemethoden basieren. Die Behandlung besteht aus mehreren aufeinander aufbauenden Phasen, nach deren Abarbeitung der Patient den fehlgebildeten Laut am Ende in allen Positionen im Wort und allen Sprechsituationen korrekt bilden kann.

Begonnen wird mit der Verbesserung der auditiven Fremd-, und Eigenwahrnehmung. Die Lautbildung der Umwelt wird zunächst auditiv sowohl mit der korrekten Lautbildung, als auch mit der eigenen verglichen. Sobald der Patient fähig ist, den Laut aus verschiedenen Kontexten heraus zu diskriminieren, ist er auch in der Lage, seine eigene Lautbildung zu korrigieren. In der Anbahnungsphase wird er angewiesen, seine Zunge in eine bestimmte Stellung zu bringen und selbstständig den isolierten Laut zu bilden (phonetische Lokalisation). Dann wird die Bildung des neu erlernten Lautes gefestigt, erst auf isolierter Lautebene, anschließend auf Silbenebene in Koartikulation und auf Wort- und Satzebene. Die letzte Phase beschreibt die Übertragung der korrekten Artikulation in die Spontansprache, d. h. Alltagssprache, wobei sein Umfeld mit einbezogen und seine Eigenkontrolle stetig verbessert wird [WZ11, S. 66ff.].

Bei phonetischen Störungen kann die Lautanbahnung durch gleichzeitige Bewegung der Hände bzw. Füße mit der Zunge unterstützt werden (bewegungsunterstützte Lautanbahnung). Die Bewegung des Körperteils berücksichtigt bei der Imitation der Zunge die Artikulationsart, den Artikulationsort und die Bewegungsrichtung der Zunge.

Gerade bei jüngeren Kindern mit hohem Leidensdruck wird die korrekte Lautbildung durch dieses Konzept erleichtert [WZ11, S. 70ff.].

Die myofunktionelle Therapie nach Kittel behandelt orofaziale Dysfunktionen, die auf ein Muskel-Ungleichgewicht und damit auf eine gestörte Schluckbewegung zurückzuführen sind. Zunächst muss die korrekte Zungenruhelage erlernt werden, da sie Laute wie [s]/[z] oder [ʃ] beeinflusst. Bei dieser liegt der vordere Teil der Zunge am Alveolarrand des Gaumens und der mittlere am harten Gaumen an. Deshalb verweist dieses Konzept häufig auf Patienten mit Sigmatismen oder Schetismen, bei denen die Ursache ein gestörtes Schlucken ist. Übungen mit Zunge und Lippen und Ansaugübungen verbessern die Koordination und Bewegungssteuerung und stärken die beteiligte Mus-

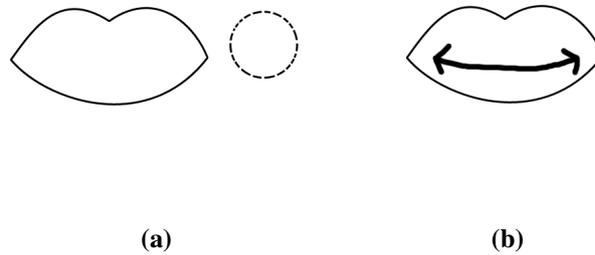


Abbildung 3.2: Bildmaterialien für myofunktionelle Übungen

kulatur. Die Abbildungen 3.2 stellen gängige Übungsmaterialien dar. Die linke Übung beschreibt das Drücken der Zunge von Innen gegen den linken Wangen, die rechte beschreibt die Bewegung der Zunge zwischen den Mundwinkeln. Das Schlucken kann dann mit Speichel, festerem Speisebrei und schließlich mit Flüssigkeiten geübt werden. Das korrekte Schluckmuster wird am Ende der Behandlung gefestigt und automatisiert [WZ11, S. 73f.]

Verbale Entwicklungsdyspraxien

Methoden für die Behandlung von Kinder mit einer VED sind im deutschsprachigen Raum u. a. VEDiT³, TAKTKIN⁴, KoArt⁵ und der Assoziationsmethode⁶, welche hier kurz beschrieben werden [Gro14, S. 185ff.].

VEDiT basiert auf dem Prinzip, möglichst viele Wahrnehmungskanäle (visuell, auditiv, taktil) verknüpft bei der Aneignung der Artikulationsbewegung mit einzubeziehen und auch als Feedback-Möglichkeit zu nutzen. Bei der Phonbildung wird ein Handzeichensystem visuell und taktil-kinästhetisch⁷ eingesetzt, welches Informationen über den Artikulationsort, -modus und Luftstromlenkung mit vermittelt. Wörter und Phrasen werden über phonemische Vereinfachungen geübt.

³s. Schulte-Mäter, A.: *Verbale Entwicklungsdyspraxie und der Therapieansatz VEDiT*, 2010

⁴s. Birner-Janusch, B.: *TAKTKIN. Ein Therapieansatz zur Behandlung sprechmotorischer Störungen wie Dysarthrien und Sprechapraxien bei Kindern*, 2006

⁵s. Stahn, C.: *KoArt. Ein Ansatz zur Therapie der Verbalen Entwicklungsdyspraxie und sprechmotorisch begründeter Aussprachestörungen*, 2010

⁶s. McGinnis, M.: *Aphasic children. Identification and training by the association method*, 1939, und Meir, R.: *McGinnis Mod.*, 2007

⁷Tast- und Berührungsempfinden und Lage- und Bewegungsempfinden

TAKTKIN ist ein Akronym des Begriffs „Taktile-Kinästhetik“ und stellt eine Erweiterung der PROMPT-Methode (*Prompts for Restructuring Oral Muscular Phonetic Targets*) auf die deutsche Sprache dar. Die Sprechbewegungen werden durch den Therapeuten taktile verdeutlicht. Parameter für die Sprechmotorik sind die Kieferöffnungsweite, die Stimulation am Mundboden, um die Zungenposition zu beeinflussen, und die Stimulation der Gesichtsmuskulatur für oberflächliche Hinweisreize. Der Therapeut kann damit direkt auf die Motorik Einfluss nehmen, falls ein Bewegungsmuster nicht korrekt ausgeführt wird. Das Übungsmaterial besteht aus Wörtern mit einfachen Silben, die häufig wiederholt werden, damit die Abfolge von Lauten eingeübt werden kann [WZ11, S. 29].

Auch beim KoArt-Ansatz wird für das Kind der passende Stimulus verwendet, um mit diesem die korrekten sprechmotorischen Bewegungen schrittweise zu entwickeln. Geübt werden diese zuerst mit Einzellauten, in der nächsten Stufe mit sinnfreien Silbensequenzen und auf oberster Ebene mit Einzelwörtern und Phrasen.

Die Assoziationsmethode nach McGinnis und ihre deutsche Abwandlung „McGinnis Mod.“ nach Meir basiert auf einer Assoziation von einem gebildeten Laut mit Bildkarten bzw. seinem geschriebenen Lautsymbol. Das Mundbild des Therapeuten, der taktile-kinästhetische Eindruck und der Lautklang werden als zusätzliche Sinneskanäle miteinbezogen.

Dysarthrien

Im Gegensatz zu den Therapieansätzen der vorangegangenen Störungsbilder sind bei Dysarthrien standardisierte Verfahren oftmals unbrauchbar. Wegen zu vieler verschiedener Erscheinungsformen sind individuelle Ansätze mit spezifischen Übungen nötig, die Behandlung wird bei der Störungsquelle begonnen. Ausgleichende Maßnahmen werden getroffen, die die verbleibenden Fähigkeiten für eine Therapie gezielt einsetzen oder es werden neue Sprechmuster erarbeitet. Ebenso können elektronische Apparate wie Stimmverstärker oder Biofeedback-Geräte, die mithilfe von Elektroden die Aktivität der Zungenmuskulatur messen, zum Einsatz kommen. In den Aufgabenbereich des Therapeuten fällt es auch, Kommunikationsstrategien für den Patienten zu vermitteln [Gro14, S. 279f.].

Sprechapraxien

Auch bei Sprechapraxien hängen die sprechmotorischen Anteile der Therapiemaßnahmen vom Grad der Defizite des Klienten ab, sodass, neben der Verbesserung von sprechmotorischen Funktionen, sensorische und psychosoziale Begleitstörungen individuell berücksichtigt werden müssen.

Eine verbesserte Verständlichkeit und Natürlichkeit beim Sprechen kann durch oft wiederholende sprechmotorische Abläufe erreicht werden. Eine Aufgabenstellung dazu wäre Nachsprechen und lautes Lesen. Das Stimulusmaterial sollte sich auf verbale Übungen mit Silben, Wörter oder Phrasen statt nur Einzellauten beschränken. Weiterhin sollten therapeutische Vermittlungstechniken zur Anwendung kommen, die die sprechmotorischen Leistungen des Klienten vereinfachen, wie es beispielsweise bei der TAKTKIN-Methode der Fall ist [Gro14, S. 283f.].

Schluckstörungen

Primäres Ziel der Therapie einer Dysphagie ist das Ernähren zu gewährleisten und die tiefen Atemwege zu schützen. Weil einzelne Bewegungen wie beispielsweise in der oralen Phase willkürlich steuerbar sind, ist das Ziel der weitergehenden Therapie über sensomotorische Übungen das physiologische Schlucken partiell oder gar gänzlich wiederherzustellen. Dafür muss die Schluckmuskultur trainiert werden, welche vorbereitend stimuliert werden kann. Das ist jedoch abhängig von dem Grad der individuellen Störung. Eine kompensatorische Therapie zielt dagegen auf ein effizientes und aspirationsfreies Schlucken ab, falls keine Verbesserung erfolgt ist. Dabei werden verschiedene Schlucktechniken geübt [Gro14, S. 309f.].

Kapitel 4

Verwandte Forschungsansätze

Die Durchführung der Therapiesitzung erfordert eine geeignete Auswahl der Medien, die der Wissenvermittlung dienen, und entsprechende Methoden um den Behandlungsablauf zu organisieren und um die Beziehung zum Klienten qualitativ hochwertig zu gestalten. Das Spektrum der Einsatzmöglichkeiten von Computer-Anwendungen im logopädischen Kontext ist breit gefächert, es reicht von audiovisuellen Tools bis hin zu Anwendungen für chirurgische Eingriffe; eine Auswahl soll hier exemplarisch vorgestellt werden.

Spielerische Therapie-Methoden helfen beim kommunikativen Austausch vor allem zwischen Patienten im Vorschulalter und dem Therapeuten. Kinder fühlen sich in einer gewohnten Umgebung sicher und empfinden im Spiel die Therapie nicht als Übungssituation. Die Eltern können leichter mit in die Therapiesitzung eingebunden werden.

Digitale Lernspiele stellen ein solches Konzept dar (siehe Edutainment). Sie erlauben ein selbstständiges Üben außerhalb der Therapie und in gewohnter Umgebung, wobei sich der Lernfortschritt im Idealfall gleichzeitig entfernt überwachen lässt. Sonst schwer zu lernende Aspekte werden durch spielerische Elemente leichter vermittelbar. Inhalte können im Gegensatz zu Bildkarten exakter und dynamisch dargestellt und vor allem so oft wie nötig wiedergegeben werden. Häufig bedienen sich die Anwendungen multimodaler Elemente durch eingebettete Video- und Audioclips zur Vermittlung der Inhalte. Die Mehrheit solcher Lernprogramme und Spiele-Apps sind für die Aphasie-Therapie bestimmt, als Beispiel lassen sich u. a. *Aphasie@ware*¹, *Revivo*² oder *SpeechCare*³ anführen.

Die Methode der „sprachtherapeutischen Geschichte“ wird genutzt, um Kindern in schwierigen Situationen unbewusst alternative Lösungswege aufzuzeigen. Mithilfe von Meta-

¹Aphasie@ware, http://www.aphasiaware.de/publikationen/publikationen_cea.htm (letzter Aufruf am 03.11.2014)

²Revivo, <http://www.aphasie.com/ueber-revivo.html> (letzter Aufruf am 03.11.2014)

³SpeechCare, <http://www.speechcare.de/die-apps/erwachsene/aphasie/> (letzter Aufruf am 03.11.2014)

phern und Symbolen kann sich das Kind in der Geschichte wiedererkennen und erhält dadurch das Gefühl nicht alleine zu sein. Gleichzeitig hat sie den psychologischen Effekt, dem Kind die Angst vor verbaler Kommunikation und Misserfolgen zu nehmen. Häufig tritt die Geschichte in Form eines interaktiven Buches auf, in dem Animationen für Übungsaufgaben eingebaut sind, die von zu Hause abgerufen werden können, oder dem traditionellen Vorlesen dienen. Fingerpuppen können dazu genutzt werden, die nachzuahmenden Bewegungen vorzumachen. In [Beł13] wird eine therapeutische Geschichte für Kinder entwickelt und in die Sprachtherapie so eingebettet, dass sie als zweidimensionale Hilfe für die korrekte Aussprache des Lautes [r] im Polnischen fungiert.

Programme, die sich speziell an hörgeschädigte bzw. gehörlose Patienten richten, basieren auf der Überprüfung der Artikulation mithilfe von audiovisuellen Elementen.

Die *SpeechMaster*-Software von Koscor und Paczolay [KP06] bietet visuelle Rückmeldung für die Korrektheit der eigenen Artikulation im Ungarischen. Trainiert wird die korrekte Verknüpfung von Phonen zu Graphemen über Spracherkennung. So müssen Spiel-Elemente über Stimmgebung, Lautstärke, Tonstufen und Rhythmus visuell bewertet oder die Aussprache von Vokalen in Wörtern und Sätzen erlernt werden.

Tiefergehende Überlegungen zu audiovisuellen Anwendungen treffen Kröger et al. in [KBHM10] für die von ihnen entwickelte *SpeechTrainer*-Software⁴. Es wird beschrieben, dass ein wichtiges Mittel zum Erlernen einer Zielsprache nicht nur Audio-Vorgaben seien, sondern ebenso Sprachaufnahmen des Lernenden selbst, um sich seiner eigenen Sprachproduktion bewusst zu werden. Die relevanten Merkmale des Sprachsignals könnten visualisiert werden, z. B. in Form eines Spektrogramms. Sprachdatenbanken helfen bei der automatischen Erkennung phonetischer Fehler. Um die korrekte Artikulation zu trainieren, sollten gleichzeitig Hilfen eingesetzt werden, die in der Lage sind Sprechorgane, den Vokaltrakt sowie deren Positions- und Bewegungsmöglichkeiten zu visualisieren.

Konzepte, die mehr den wissenschaftlich-realtätsnahen denn einen spielerisch-pädagogischen Ansatz beanspruchen, finden sich in dreidimensionalen, biomechanischen Modellen für medizinische Simulationen.

Die Biomechanik befasst sich mit den Wechselwirkungen von Kräften auf biologischen Strukturen und die daraus resultierenden Bewegungen, Belastungen und Beanspruchungen, die untersucht und beurteilt werden. Dafür müssen reale, physikalische Zusammenhänge und Größen in ein vereinfachtes Modell überführt werden, welches in einer Simulation Anwendung findet. Viele Ansätze hierzu machen sich die Finite-Element-Methode (FEM) als grundlegendes Hilfsmittel zunutze, bei der eine räumliche Struktur – ein Modell mit endlichen Teilelementen – einer komplexen Struktur durch Unterteilung in kleinere Elemente angenähert wird. Für jedes dieser Elemente werden dann

⁴SpeechTrainer, <http://www.speechtrainer.de/> (letzter Aufruf am 03.11.2014)

mechanische Gleichungen individuell gelöst, um darauf Rückschlüsse auf die globale Verformung der Struktur zu ziehen. FEM kommt insbesondere dann zum Einsatz, wenn das Deformationsverhalten von Körperteilen wie z.B. der menschlichen Zunge beschrieben werden soll. Andere Techniken basieren auf Feder-Masse-Systemen um die Mechanik von Weichgewebe zu repräsentieren. Ein Wissen über die Muskelphysiologie ist für den medizinischen Einsatz bei beiden Varianten unabdingbar. Muskelmodelle dienen der Berechnung von Muskelkontraktionen, für die Bewegungsanalyse sowie für die Bestimmung von Belastungen.

Mit dem Thema der biomechanischen Modellbildung für die Sprachforschung befassen sich eingehend Perrier et al. in [PPB⁺11]. Die von ihnen verwendete 3D-Geometrie der Artikulationsorgane basiert auf exakten anatomischen Daten aus dem *Visible Human Project*⁵, die mithilfe von MRI- und CT-Scans angepasst wurden.

Einen umfangreichen Überblick über die Methodik und zahlreiche Verweise auf Modelle der menschlichen Zunge gibt Wilhelms-Tricarico in dem Vortrag „From Muscle Models to Tongue Models (And Back)“ [WT06]. Wilhelms-Tricarico und Perkell stellen schon in ihrer frühen Arbeit [WTP97] ein approximiertes Modell des Vokaltrakts vor, das versucht Sprechbewegungen mithilfe der FEM zu simulieren, und dabei so viel Wissen wie möglich über die Muskelphysiologie mit einzubeziehen.

Biomechanische Modelle können auch als Hilfsmittel bei chirurgischen Eingriffen und im Sinne einer prädiktiven Medizin eingesetzt werden. So stellen Rodrigues et al. [RGC98] eine Simulation über das Verhalten der oberen Atemwege während einer Laryngoskopie (Kehlkopfspiegelung) vor. Die Visualisierung soll es dem Anästhesisten erlauben, Voraussagen über den Eingriff zu treffen und schwierige Situationen zu trainieren. Hauptbestandteil der Modellierung ist bei diesem Ansatz das physische Verhalten der Zunge, da sie für die Spiegelung mithilfe eines Zungenhalter verschoben und eingengt werden muss. Nicht nur die Antwort des Gewebes auf das Instrument müssen repräsentiert werden, sondern auch die Gestalt der Zunge.

In einer Simulation von Buchaillard et al. [BBPP07] wird ebenso das Modell der Zunge als Teil der Mundhöhle in Form einer Finite-Elemente-Struktur verwendet, um die Auswirkungen einer Hemiglossektomie (asymmetrische Entfernung des Zungenkörpers) auf die Zungenbeweglichkeit und damit auch auf die Sprachproduktion vorhersagen zu können. Das fehlende Gewebe nach dem operativen Eingriff korrespondiert mit fehlenden Elementen im Modell.

Im Kontext des computeranimierten Sprechens bietet das parametrische Zungenmodell von King und Parent [KP01] die Möglichkeit, die Form der Zunge und ihre Bewegung bei der Bildung von englischen Vokalen und Konsonanten zu animieren. Das Modell

⁵Visible Human Project, http://www.nlm.nih.gov/research/visible/visible_human.html (letzter Aufruf am 03.11.2014)

besteht aus einer Oberfläche von B-Splines Patches, eine Parametrisierung bestimmt den Raum der möglichen Zungenstellungen und verformt die Geometrie.

Ein parametrisches Modell von Ilie et al. mit automatisch bestimmten virtuellen Knochen erlaubt es, korrespondierende Segmente der Zunge über fünf Kontrollpunkte zu manipulieren [INS12]. Die virtuelle Skelett-Struktur orientiert sich dabei an den geometrischen Merkmalen der Zunge. Die Transformation dieser geschieht mithilfe eines mathematischen Modells, um den Einflussbereich des virtuellen Knochens auf die umgebenden Meshpunkte zu ermitteln. Das Mesh der Zunge ist dabei manuell modelliert.

Bei der Konzeption der vorliegenden Arbeit spielten der oben bereits erwähnte *Speech-Trainer* und die Konzepte mit virtuellen Sprechern eine große Rolle. Beide Ansätze werden im folgenden Kapitel aufgegriffen.

Kapitel 5

Konzept

In diesem Kapitel wird die grundlegende Idee und die damit verknüpften Anforderungen der vorliegenden Arbeit beschrieben. Weiterhin werden die einzelnen Arbeitsschritte und die Auswahl der Werkzeuge erläutert, welche in einem ersten Prototypen resultieren.

5.1 Beschreibung

Das Konzept beruht auf der Beobachtung, dass in der Sprechtherapie in der Regel keine Computer-Anwendungen zur Unterstützung eingesetzt werden, die die Lautbildung visualisieren, obwohl es Logopäden in diesem Kontext durchaus als sinnvoll erachten würden (siehe Kapitel 1.1). Ziel dieser Arbeit ist demnach die Entwicklung eines therapeutischen Werkzeuges für die Vermittlung von Artikulationsbewegungen der deutschen Sprache und des Schluckvorgangs, welches ein dreidimensionales Modell der oro-pharyngo-laryngealen Region integriert. Der Sprechtherapeut soll das Modell um die eigene Achse drehen und vergrößern können, damit es aus jeder möglichen Perspektive eine freie Sicht auf die Artikulationsorgane während der Lautbildung und des Schluckens gibt. Die Bewegungen der Artikulationsorgane sollen anatomisch exakt dargestellt werden.

Die umzusetzende Funktionalität ähnelt den Ideen der audio-visuellen Anwendung *SpeechTrainer* und der virtuellen Tutoren *Baldi* und *ARTUR*, weshalb diese in einigen Aspekten als Vergleich herangezogen werden können.

Kröger et al. stellen in [KBHM10] die *SpeechTrainer*-Software für die deutsche Sprache vor, die visuelle und zugleich auditive Mittel nutzen, um die phonetischen Fähigkeiten des Patienten zu verbessern. Der zu erlernende Ziellaut wird aus einer Liste deutscher Vokale, Nasale, Frikative oder Laterale ausgewählt und der Patient muss sich mithilfe einer Spracheingabe diesem spektralen Umfang annähern. Der Ziellaut wird akustisch

ausgegeben, sodass es dem Lernenden möglich ist, beide Spektren der Sprachsignale in einem Spektrogramm o. ä. bezüglich Intonation oder Lautlänge zu vergleichen. Der Laut des Patienten wird nicht nur akustisch, sondern auch artikulationsbezogen analysiert, d.h. eine hier vorgestellte visuelle Hilfe bietet eine zweidimensionale sagittale Sicht auf den Vokaltrakt und die Artikulationsorgane während der eigenen Lautproduktion. Für die Visualisierung wird der zum eingegebenen Laut akustisch nächst gelegene in einer zweidimensionalen Ansicht angezeigt.

Der virtuelle Tutor *Baldi* von Massaro [Mas06] stellt einen Assistenten zur Sprachtherapie von tauben und schwerhörigen Kindern dar. Der Ansatz ist hier ebenso ein anatomisch realistisches Modell zu präsentieren, an dem die Sprachproduktion visualisiert wird, die Zielsetzung der Arbeit liegt in der Sprachsynthese. Der Tutor gibt Übungen zum Vokabular vor und bietet gleichzeitig beim Sprechen eine dreidimensionale Innenansicht der Mundhöhle. Für die Darstellung eines Phons werden Zielwerte wie Kieferstellung, Mundweite etc. gesetzt, Äußerungen werden als String von Phonemen repräsentiert.

Mit einem virtuellen Sprecher für die Artikulation der schwedischen Sprache beschäftigt sich das Projekt *ARTUR*¹ von Engwall et al., das sich ebenfalls an hörgeschädigte Kinder richtet. In [EBOK06] werden anhand des Projekts Anforderungen an ein audiovisuelles System im Allgemeinen gestellt. Ein wichtiges Ziel sei u. a. eine Erkennung von falscher Aussprache, um dem Patienten angemessene Rückmeldung geben zu können. Dafür werden statistische Methoden der Spracherkennung verwendet sowie die Extrahierung visueller Informationen aus dem mit Video aufgenommenen Gesicht des Patienten, um die Erkennung der falschen Aussprache zu verbessern. Mithilfe einer „artikulatorischen Umkehrung“ kann anschließend über die auditive und visuelle Eingabe die Lautbildung des Benutzer annähernd rekonstruiert und in einem dreidimensionalen Artikulationsmodell dargestellt werden, welches eine Innenansicht des Vokaltrakts bietet.

Im Vergleich zu solchen Konzepten ist das vorrangige Bestreben dieser Arbeit ein therapeutisches und kein patientenseitiges Hilfsmittel zu entwickeln, deshalb sollte so oft wie möglich die Expertenmeinung von Logopäden in den Umsetzungsprozess der Anwendung mit einfließen. Vor der Umsetzung der Arbeit geschah dies über Hospitation in zwei logopädischen Einrichtungen, während der Umsetzung auch über regelmäßiges Feedback von Seiten der Experten. Langfristiges Ziel dieser Arbeit ist es demnach, eine solche Anwendung in den Therapie-Ablauf von Logopäden zu integrieren und eine Ergänzung zu bisherigen Materialien zu bieten.

¹ARTUR, <http://www.speech.kth.se/multimodal/ARTUR/> (letzter Aufruf am 03.11.2014)

5.2 Anforderungen

Im Folgenden werden die therapeutischen Anforderungen an die Anwendung, die sich aus der ersten Expertenumfrage ergaben, betrachtet. Auf Grundlage dessen und unter Einbeziehung der Zielsetzung dieser Arbeit konnte eine Auswahl an Kernfunktionen getroffen werden, die im Prototypen implementiert werden sollten.

5.2.1 Therapeutische Anforderungen

Um eine solche Anwendung in die Übungen der Sprechtherapie einzugliedern, muss sie zunächst den Anforderungen aus logopädischer Sicht genügen. Die Expertenmeinungen der zwölf Logopäden aus Tabelle 1.1 wurden eingeholt, damit ein erster Prototyp erstellt werden konnte. Gefragt wurde deshalb nach allgemeinen Anforderungen für eine Computer-Anwendung in der Sprechtherapie.

Aus therapeutischer Betrachtung ergibt sich, dass die Anwendung aufgrund von zeitlich knapp bemessenen Therapie-Einheiten ohne lange Vorbereitungszeit einsetzbar sein sollte, damit sie als unterstützendes Material hinzugezogen werden kann. Auch die Bedienung sollte deshalb intuitiv sein. Falls eine Benutzung der Anwendung seitens des Patienten mit eingeschlossen würde, müsse die Handhabung je nach Störungsbild patientengerecht verlaufen. Vor allem Patienten mit Schlaganfall würden einer vereinfachten Bedienung beispielsweise mit einer Hand oder eine Steuerung über Eye-Tracking bedürfen. Die Anwendung sollte aber nicht nur für Erwachsene komfortabel bedienbar sein, sondern auch für Kinder. Für diese müsse dann ein motivierender Faktor in der Anwendung geschaffen werden. Wichtigste Bedingung an die Anwendung sei jedoch die physiologische Korrektheit bei allen vorzuführenden Bewegungen, da sie sonst keinen Mehrwert zu einfachen Bildkarten biete.

Als technische Anforderung wird die Möglichkeit genannt, das Abspielen von Sprechbewegungen verlangsamen zu können, damit Patienten nicht abgehängt würden und Übungen direkt mitmachen könnten. Gerade bei der Präsentation von hintereinander gereihten Lauten in einem Wort (Koartikulation) sei ein manuelles „Durchscrollen“ durch das Wort optimal, um so alle Bewegungen wiederholt und im langsamen Tempo ansehen zu können. Sinnvoll wäre es zudem, wenn man auf einen abgespeicherten „Lernstand“ des Patienten zurückgreifen könne. Ein Ausdrucken einer konkreten Stellung der Artikulationsorgane ergänze die gewöhnlichen Bildmaterialien und biete gleichzeitig dem Patienten die Möglichkeit einer Rückmeldung über seine Schwierigkeiten. Damit lasse sich das Material auch einfach mit nach Hause zum weiteren Üben nehmen.

Andere Argumente nehmen Bezug auf ökonomische Aspekte: der überwiegende Teil der logopädischen Einrichtungen erhalte kaum finanzielle Mittel für solche Anwendungen, die Anforderung sei deshalb als ein wirtschaftlich vertretbares Produkt zu entwerfen. Zusätzlich wurde erwähnt, dass die Anwendung auf Tablet und Smartphone nutzbar sein könne.

5.2.2 Funktionale Anforderungen

Auf Grundlage der allgemeinen therapeutischen Anforderungen und der Zielsetzung dieser Arbeit wurden funktionale Anforderungen für einen ersten Prototypen in einem Mock-Up (Abb. 5.1) festgehalten:

- Der Therapeut soll den Blickpunkt auf das 3D-Modell frei wählen können. Dieses muss sich demnach in alle Raumrichtungen drehen lassen und ein Heranzoomen erlauben
- Teile des Modells wie Muskeln, Mundhöhle und Unterkiefer sollen sich in der 3D-Ansicht ein- und ausblenden lassen
- Der Therapeut kann aus einer Liste von vorgegebenen Phonen dasjenige auswählen, dessen Artikulationsbewegung in Form einer Animation vorgespielt werden soll
- Die Zungenruhelage soll auf die gleiche Art ausgewählt und vorgeführt werden können
- Die Animation soll vom Therapeuten gestoppt und wiederholt werden können
- Die Bewegungen der Zunge müssen während der Bildung eines Lautes anatomisch korrekt visualisiert sein
- Der Therapeut soll die Möglichkeit haben, die Zunge anatomisch korrekt über ihre drei extrinsischen Muskeln bewegen zu können.

5.3 Herangehensweise

Nachdem die Anforderungen und grundlegende Funktionalität bestimmt waren, wurde die Anwendung iterativ umgesetzt; zunächst in einem ersten Prototypen, welcher anschließend mithilfe von Logopäden über Fragebögen beurteilt wurde. Die daraus gewonnenen Erkenntnisse flossen dann in die Implementierung des finalen Prototypen mit ein.

Zu Anfang musste für die Umsetzung der Anwendung eine Simulationsumgebung und ein Visualisierungsmodell gewählt werden. Die Werkzeugauswahl wird im Folgenden näher erläutert.

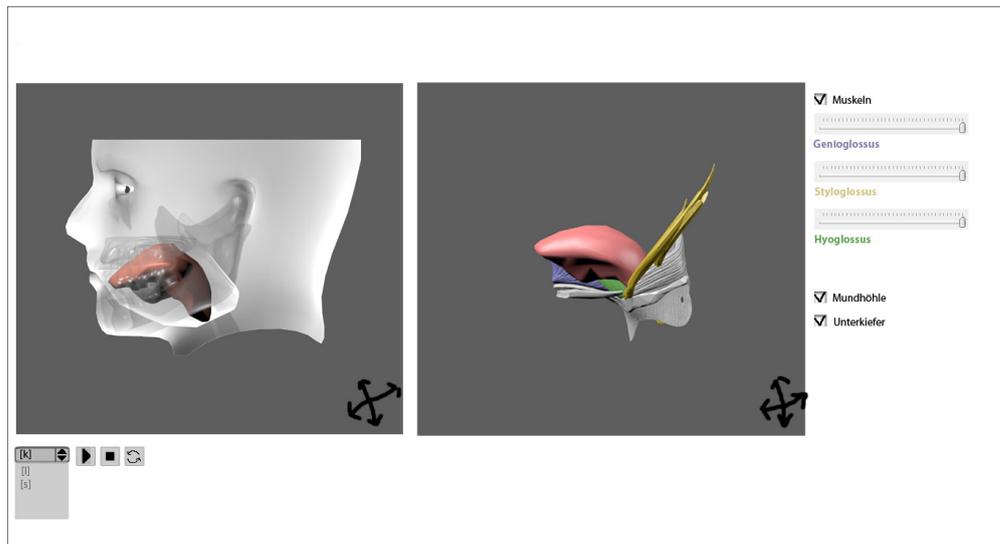


Abbildung 5.1: Ein erstes Mock-Up der Anwendung. Die Bedienoberfläche bietet zwei Ansichten des 3D-Modells, eine Auswahlliste von Lauten, Schieberegler um die Zunge nach ihren Muskeln zu bewegen und Kontrollkästchen um Modellteile ein- und ausblenden zu lassen. Der überwiegende Teil der Bedienoberfläche wurde für den ersten Prototypen übernommen.

5.3.1 Visualisierung

Als 3D-Grafikumgebung wurde die Modellierungssoftware Blender² gewählt, da diese bereits über einen großen Funktionsumfang verfügt, der bei dieser Arbeit von Nutzen war.

Zum einen bietet Blender als Modellierungstool von Haus aus das 3D-Ansichtsfenster, das mithilfe der Maus gesteuert werden kann. So war es nicht nötig, ein eigenes Framework für diese Arbeit zu entwickeln, das Modelle importiert und Mausinteraktionen zulässt. Zusätzlich integriert Blender zwei eigene Render-Engines, sodass die Visualisierungsmethoden in Blender dementsprechend umfangreich sind. Daher können die Renderingeigenschaften einer Szene vom Benutzer schnell an die eigenen Bedürfnisse angepasst werden. Dem Logopäden kann dadurch eine konsistente Arbeitsweise mit gleichbleibender Visualisierung geboten werden, die kein Wechsel zwischen verschiedenen Übungsmaterialien erfordert.

Zum anderen verfügt Blender über eine Python-Programmierschnittstelle, mit deren Hilfe auf Daten des User-Interfaces und der Modelle zugegriffen werden kann um diese zur Laufzeit zu manipulieren. Über ein Script können so zum einen einzelne Vertices eines Modells angesteuert werden, zum anderen aber auch Elemente wie Menüs oder Panels zur Bedienoberfläche hinzugefügt werden. Die Schnittstelle bietet hierzu verschiedene sog. Application Modules an, um Objektdaten, ihre Polygonnetze, Typen,

²Blender, <http://www.blender.org/> (letzter Aufruf am 04.11.2014)

Operatoren oder den aktuellen Kontext in Blender abzufragen. Vor allem im Entwicklungsprozess der Arbeit war die Möglichkeit, jederzeit Funktionalität unkompliziert hinzuschalten, erweitern und ausblenden zu können, von Vorteil.

Um dem Therapeuten für die Bedienung eine Einarbeitung in die Blender-Oberfläche zu ersparen, sollte das Interface auf das Wesentliche beschränkt werden. So wird der Fokus wie gewünscht auf die Visualisierung gelegt und die Frage der Bedienbarkeit für diese Arbeit in den Hintergrund gerückt. Die Interaktion sollte neben der Maus über einfache Bedienelemente gewährleistet werden. Viele Panels und Menüs der GUI lassen sich manuell ausblenden oder mithilfe der Application Modules überschreiben. Für diesen Zweck konnte die Blender-Oberfläche auf drei Regionentypen reduziert werden. Diese Oberfläche wurde in beiden Prototypen eingesetzt.

5.3.2 Modelle

Die in dieser Arbeit verwendeten Modelle für die Artikulationsorgane stammen hauptsächlich von der BodyParts3D³-Webseite, die vom Database Center for Life Science der Universität Tokio betrieben wird. Sie stellen dort Modelle des menschlichen Körpers zum Herunterladen im OBJ-Format bereit. Alle zusammengehörigen Organe liegen gruppiert in einem großen Körpermodell vor, weswegen die Proportionen der einzelnen Objekte nicht weiter bearbeitet werden mussten. Zusätzlich sind sie jeweils mit anatomischen Labels versehen, sodass ein schneller Zugriff auf die notwendigen Modelle der oro-pharyngo-laryngealen Region erfolgen konnte.

Ein beispielhaftes, anatomisch vereinfachtes Modell eines menschlichen Kopfes erstellt Karan Shah in einem Modellierungs-Tutorial⁴. Die Vorlage verfügt bereits über grundlegende Mimiken und das Polygonnetz ist im Vergleich zum BodyParts3D-Kopf deutlich vereinfacht. Bei der Lautbildung sind vor allem die Lippen und der Kiefer die wichtigsten Akteure, für Animationen mussten sie leicht transformiert werden können. Wegen seiner einfachen und symmetrisch aufgebauten Polygonstruktur schien dieses Kopfmodell für diese Anwendung ausreichend.

Für das anatomische Gesamt-Modell mussten das Modell des Kopfes und das der Artikulationsorgane zusammengeführt und die Größenverhältnisse untereinander angepasst werden. Dafür wurde das komplette Modell mit gängigen anatomischen Bildmaterialien über den Bau des Ansatzrohres im Sagittalschnitt verglichen. Alle Modelle wurden abschließend in der Anzahl ihrer Vertices so reduziert, dass ihr Aussehen beibehalten wurde, und ihre Oberflächen in Blender geglättet. Für den Bewegungsablauf der Laut-

³BodyParts3D©, lizenziert unter CC BY-SA 2.1 JP, <http://lifesciencedb.jp/bp3d/> (letzter Aufruf am 04.11.2014)

⁴Tutorial: Create a Facial Animation Setup in Blender, <http://cgi.tutsplus.com/tutorials/create-a-facial-animation-setup-in-blender-part-1--cg-32251> (letzter Aufruf am 04.11.2014)

bildung und den Schluckvorgang musste jedes Modell anschließend manuell so animiert werden, dass ihr Zusammenspiel anatomisch korrekt ausfiel.

Die Animation eines Modells wird im Allgemeinen über sog. Shape Keys in Blender umgesetzt, von denen mehrere pro Objekt angelegt werden können. Ein Shape Key repräsentiert eine Zieltransformation der Mesh-Vertices des Objekts und damit die Form des Objekts. Über den Wert des Shape Keys, der für gewöhnlich zwischen 0 und 1 liegt, kann sein Einfluss pro Frame festgelegt werden. Eine Animation entspricht dann der räumlichen und zeitlichen Interpolation zwischen einzelnen Shape Keys. Es ist dabei möglich, für jeden Shape Key einen eigenen Interpolationsmodus zu wählen, standardmäßig wird in Blender zwischen den einzelnen Frames eines Keys eine Bézier-Interpolation angewandt.

Kapitel 6

Implementierung

Nachdem das Grundgerüst der Simulationsumgebung feststand, wurde das integrierte Gesamt-Modell manuell für alle Lautbewegungen animiert. Im zweiten Prototypen wurden die neu hinzugekommenen Modelle der pharyngo-laryngealen Region angepasst und für den Schluckvorgang zusätzlich animiert. Das Zungenmodell wurde verfeinert und die Animationen der Lautbildung nach der Vorlage von Fiukowski [Fiu10] anatomisch verbessert. Die Visualisierung eines Phonationsstrom benötigte ein Partikelsystem, das für jeden Laut manuell gesetzt werden musste. Die Implementierung der Bedienoberfläche und alle damit verknüpften Funktionalitäten, wie das Setzen von Keyframes, erfolgte über die Blender-Python-Schnittstelle.

Folgende Abschnitte beschreiben die konkrete Umsetzung der im vorherigen Kapitel geforderte Funktionalität im ersten Prototypen und die daran angeknüpfende logopädische Beurteilung. Anschließend wird die Implementierung des finalen Prototypen unter Berücksichtigung der Verbesserungsvorschläge erläutert.

6.1 Umsetzung des ersten Prototypen

Für die Abbildung eines Lautes benötigt jedes am Laut beteiligte Organmodell bestimmte Transformationen, die beim Abspielen der Animation angesteuert werden. Diese Zielpositionen sind für jedes Modellteil in Form von Shape Keys festgehalten. So ist beispielsweise für die Umsetzung des Lautes [l] der Kiefer ein wenig geöffnet, der vordere Zungenrand liegt an den palatinalen Flächen der oberen Zähne an und der Zungenrücken ist flach gewölbt. Um jeden Ziellaut möglichst anatomisch korrekt zu bilden, war ein Abgleich mit Bildmaterialien notwendig.

Die Verformung der Zunge findet über einen mit Shape Keys versehenen SimpleDeform-Modifiers statt, der sich quaderförmig um das Zungenobjekt legt und dessen Ursprung in einem sog. leeren Objekt am Zungenrund liegt (Abb. 6.1a). Die Verformung der Zunge wird damit über den Objektmodifikator kontrolliert, ohne dass das Mesh der Zunge

selbst transformiert werden muss. Sobald die Ziel-Zungenstellung darüber gesetzt ist, wird sie in ebenso einem Shape Key festgehalten. Zusätzlich kann die Zunge über festgelegte Vertex-Gruppen des Modifikators transformiert werden (Abb. 6.1b). So können Zungenspitze, -Körper und -Wurzel über jeweils eine Vertex-Gruppe manipuliert werden. Die Muskeln der Zunge erhalten jeweils MeshDeform-Modifier, die ebenso an den SimpleDeform-Modifier der Zunge gebunden werden und sich bei jeder Bewegung der Zunge entsprechend mitverformen.

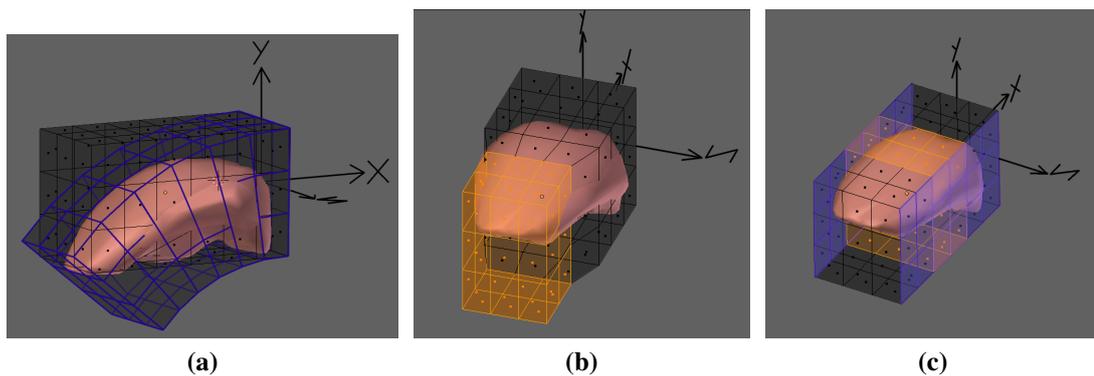


Abbildung 6.1: Zungenmodell mit SimpleDeform-Modifier und der Zungenspitze als Vertex-Gruppe. (a) Der SimpleDeform-Modifier wird um 30 Grad gekrümmt (blaues Gitternetz), (b) Die ausgewählte Vertex-Gruppe (orangenes Gitternetz) schmälet die Zungenspitze, (c) Angelegte Vertex-Gruppe für Zungenkörper (orange) und Seitenränder (blau)

Die vom Therapeuten zu bedienende grafische Oberfläche konzentriert sich auf drei Bereiche der Blender-Oberfläche, die die Funktionalität des im vorherigen Kapitel erarbeiteten Mock-Ups einbinden (Abb. 6.2c).

Das Kopf-Modell ist im 3D-View-Fenster (Abb. 6.2c (3)) visualisiert. Es stellt das Hauptfenster für die Ansicht der Modelle dar. Da der Schluckvorgang im ersten Prototypen noch nicht realisiert ist, beschränken sich die Modelle auf die der Artikulationsorgane. So stellen der Kopf inklusive Mundhöhle, Oberkiefer, Unterkiefer, die Zunge, die Muskeln *M. styloglossus* (gelb), *M. genioglossus* (blau), *M. hyoglossus* (grün) und das Zungenbein jeweils eigene Objekte dar, die auf verschiedenen Sichtbarkeits-Ebenen liegen, welche sich separat ein- und ausblenden lassen. In Abb. 6.2a und 6.2b ist das gesamte Kopfmodell im Ruhezustand aus zwei verschiedenen Ansichten dargestellt. Mund mit Lippen und Kiefer sind geschlossen und die Zunge nimmt eine mittlere Position ein, sodass die Zungenspitze locker an die vorderen Frontzähne stößt. Eine Mikrofon-Funktionalität ist im Header (Abb. 6.2c (1)) des 3D-View-Fensters eingebettet und das UI-Panel (Abb. 6.2c (2)) beinhaltet die übrigen Funktionen in Form eigens definierter,

aufklappbarer Panels. Der Therapeut kann dort interaktiv Teile des Modells ein- und ausblenden, die Zunge zu ihren Muskeln bewegen sowie Laute vorführen lassen.

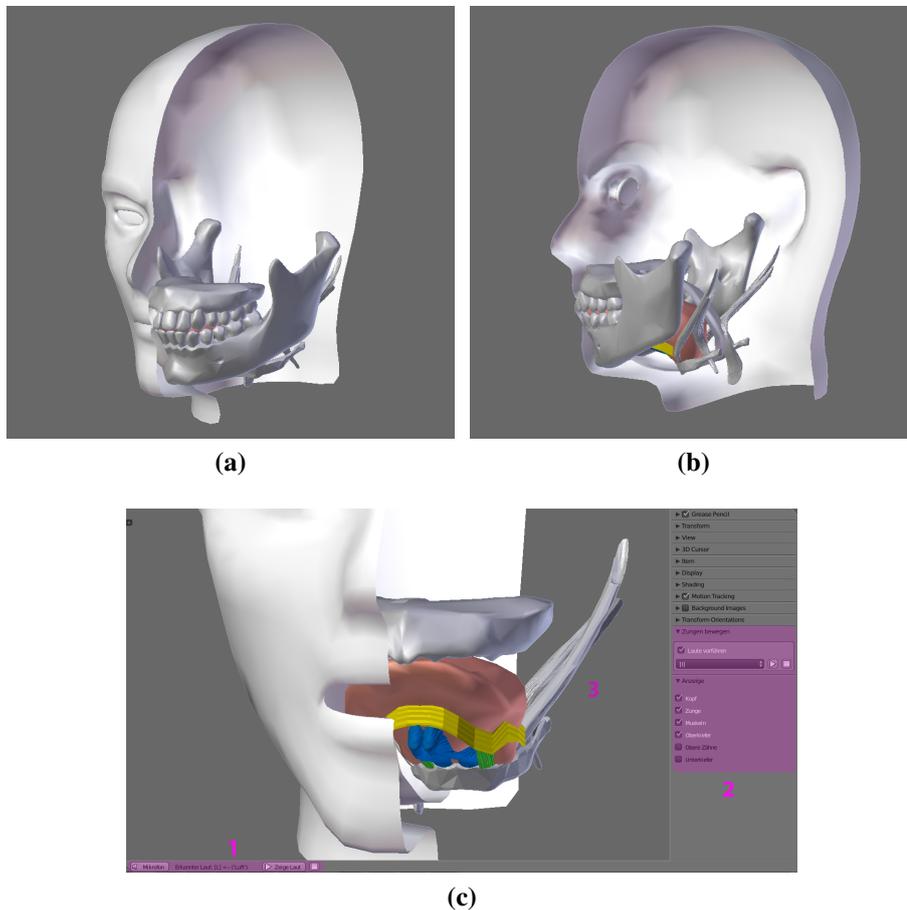


Abbildung 6.2: (a), (b) Vollständiges Modell im Ruhezustand, (c) Übersicht über das User-Interface, das auf drei Bereiche reduziert wurde. Beim ersten Prototypen findet sich Funktionalität im Header (1) und im rechten Panel (2). Das Modell lässt sich im Hauptfenster (3) drehen und bewegen.

6.1.1 Zunge bewegen

Der Therapeut kann Bewegungen der Zunge auf zwei verschiedene Arten ausführen. Er kann eine Animation für einen ausgewählten Phon abspielen lassen oder die Zunge anhand ihrer extrinsischen Muskeln manuell bewegen. Letzteres ermöglicht ihm die Zunge in eine frei wählbare Zielstellung zu transformieren und damit Zungenübungen

vorzuzeigen.

Bei aktiviertem Kontrollkästchen für das Vorführen von Lauten erscheint ein entsprechendes Menü. Aus einer Liste lässt sich der gewünschte Laut auswählen. Die Liste umfasst acht unterschiedliche Laute (Abb. 6.3a und 6.3b), von denen sich fünf vorführen lassen, da im ersten Prototypen zunächst nicht zwischen Lenis- und Fortis-Verschlusslauten unterschieden wird. Weil Zungeneinstellung und der Abschluss des Gaumens bei einem Fortis- und Lenis-Lauts gleich sind, werden sie deshalb nur als einen Menüpunkt aufgeführt (z. B. [t]/[d]).

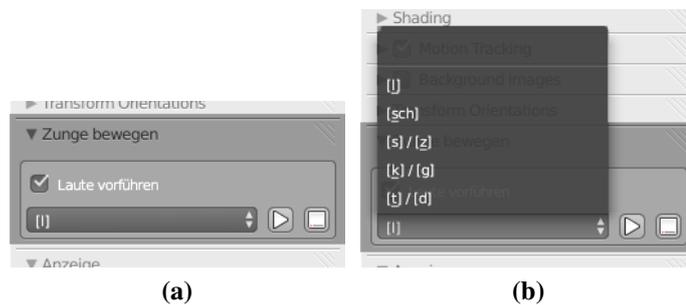


Abbildung 6.3: Menü-Bereiche des UI-Panels im Detail. (a) Laute vorführen lassen, (b) Die aufgeklappte Phon-Liste

Mithilfe eines ‚Play‘- und ‚Stop‘-Buttons lässt sich die abzuspielende Animation steuern. Sie wird in Echtzeit im 3D-View-Fenster von Blender abgespielt. Dadurch ist es dem Benutzer möglich, während des Vorspielens der Animation das Modell weiterhin zu bewegen und zu drehen. Die Anzahl der Vorschau-Frames wurde dabei für alle Animationen auf 60 festgelegt, um die Dauer der Vorführung bei ca. 2 bis 3 Sekunden zu belassen. Sobald der ‚Play‘-Button gedrückt wird, wird der gewählte Laut bzw. das gewählte Lautpaar in der Liste abgefragt. Wegen der geringen Anzahl an vorgegebenen Lauten im ersten Prototypen ist jeder Laut mit einer eigenen Funktion verknüpft, die dann die entsprechenden Keyframes für die Animation des Lautes setzt. Für jeden Laut bzw. jedes Lautpaar können darüber individuelle Shape Key-Werte verwendet werden. Für beispielsweise den Laut [l] befindet sich das Modell zu Beginn der Animation im Ruhezustand, d. h. die Shape Keys des Mundes, des Kiefers und des Zungen-Modifiers werden zum Frame 0 auf den Wert 0.0 gesetzt. Für die Darstellung des Lautes wird der Mund und der Kiefer zur Hälfte geöffnet und die Zunge nimmt vollständig die Zielposition für den Laut [l] ein. Diese Werte werden im Frame 60 gesetzt, sodass beim Abspielen der Animation zwischen den Anfangs- und den Endwerten interpoliert werden kann.

Sobald das Kontrollkästchen für das Vorführen von Lauten deaktiviert wird, öffnet sich ein neues Menü mit mehreren Schiebereglern (Abb. 6.4). Mit diesen hat der Therapeut

die Möglichkeit, die Zunge „frei“ bewegen zu können. Der Mund lässt sich öffnen, wenn der erste Schieberegler im Intervall $[0, 1]$ mit der Maus verschoben wird. Jeder Wert dieses Schiebereglers wird dann abgebildet auf die Shape Key-Werte im Intervall $[0, 1]$ von Mund, Unterkiefer und vom Modifier der Zunge.



Abbildung 6.4: Menü-Bereich um die Zunge entsprechend der Muskeln zu bewegen

Für die freie Bewegung der Zunge stehen jedem Muskel zwei Schieberegler zur Verfügung, die seine anatomische Kontraktion abbilden (vgl. Kapitel 2.1). Der Slider des Genioglossus' lässt so die Zunge nach vorne strecken oder den Zungenkörper nach unten wölben. Entgegengesetzte Bewegungen werden durch ein verändertes Intervall des Sliders ausgeführt. So kann ein Schieberegler des Hyoglossus' nach links im Intervall $[-1, 0]$ oder rechts im Intervall $[0, 1]$ verschoben werden, deren Werte wieder auf die Shape Key-Werte des Zungen-Modifiers im Intervall $[0, 1]$ abgebildet werden. Die Zunge wird dadurch nach links oder rechts um die lokale X-Achse gekippt, wie in Abbildung 6.5b zu sehen.

Um eine Verbindung direkt zum Modell herstellen zu können, welcher Slider welche Modellbewegung hervorruft, ist die Muskelbezeichnung zusätzlich farblich entsprechend seines Modellmaterials codiert. Es wird nicht geprüft, ob der Mund vor der Bewegung der Zunge geöffnet wurde, die Zunge kann auch bei geschlossenem Mund im Mundraum bewegt werden. Es hier jedoch keine Kollisionserkennung implementiert, die die Zunge davon abhält, an die Grenzen der Mundhöhle zu stoßen.

6.1.2 Ein- und Ausblenden von Objekten

Mithilfe der Kontrollkästchen im Anzeige-Panel (Abb. 6.6) lassen sich Teile des Modells wie Kopf, Zunge, Muskeln, Ober- und Unterkiefer ein- und ausblenden. Intern werden dadurch die Sichtbarkeits-Ebenen der Modelle, in der sie liegen, aktiviert oder deaktiviert. Dies ermöglicht das nähere Betrachten der Objekte, ohne dass sie von anderen Objekten verdeckt werden. Wenn der Oberkiefer eingeblendet ist, können zusätzlich

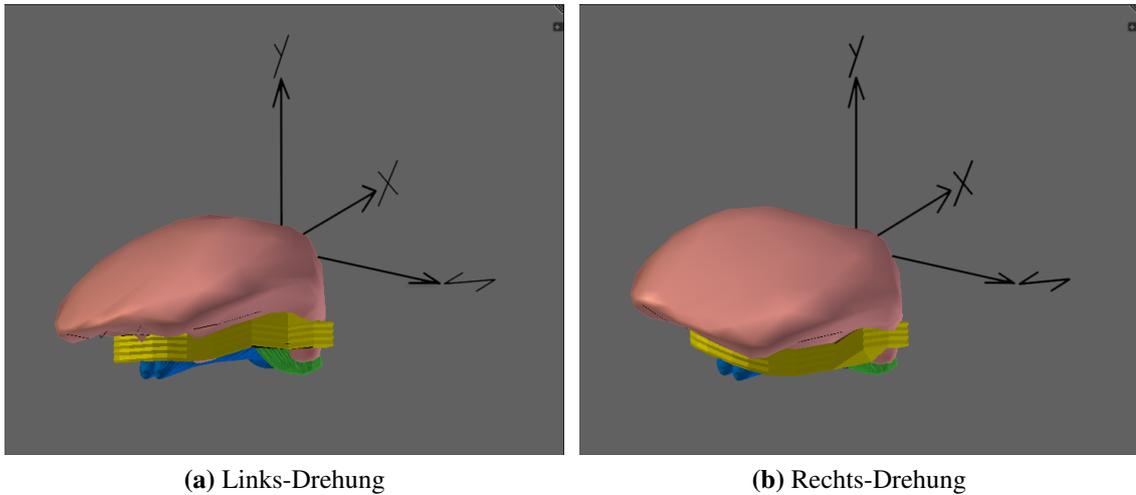


Abbildung 6.5: Der Muskel Hyoglossus (grün) lässt die Zunge links und rechts um die lokale X-Achse kippen

die oberen Zähne als Ganzes ein- und ausgeblendet werden. Die ist vor allem wichtig, um während der Lautbildung zu sehen, ob die Zungenspitze am oberen Zahndamm oder der Zungenrücken am Gaumen anliegt.

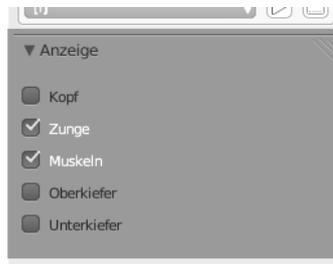


Abbildung 6.6: Das Anzeige-Panel erlaubt es, Modellteile ein- und ausblenden zu lassen.

6.1.3 Mikrofon-Anbindung

Die Mikrofon-Anbindung ermöglicht dem Patienten ein Wort einzugeben, um die Artikulationsbewegung des ersten Buchstabens als Animation abzuspielen.

Mit Drücken des Mikrofon-Buttons im Header ((1) im Interface) wird innerhalb einer vordefinierten Aufnahmezeit übertragene Sprache aufgenommen und als WAV-Datei abgelegt. Das Gesprochene kann damit jederzeit wieder abgespielt werden und der Pa-

tient erhält indirekte Rückmeldung über seine Artikulation. Eine Programmbibliothek¹ zur Spracherkennung, welche intern auf Funktionen der Spracherkennung von Google zurückgreift, wird dann verwendet, um die Audioinformation der WAV-Datei zu extrahieren und zu transkribieren. Dafür ist eine Internet-Verbindung erforderlich. Falls das Gesprochene erkannt wurde, wird es im Header angezeigt und der erste Buchstabe als vorzuführenden Laut ausgewählt. Ein ‚Play‘- und ‚Stop‘-Button erscheinen dann, um den Laut abzuspielen. Das 3D-View-Fenster von Abbildung 6.2c zeigt das Modell bei der Vorführung des Lautes [l], nachdem das Wort ‚Luft‘ von der Programmierschnittstelle erfolgreich erkannt und der erste Buchstabe des Wortes als abzuspielendes Phon automatisch selektiert wurde. Falls keine Eingabe erkannt wurde, erscheint eine Fehlerausgabe und die Eingabe kann daraufhin wiederholt werden.

6.2 Beurteilung des ersten Prototypen

Die zwölf Logopäden aus Tabelle 1.1 erhielten eine kurze Einführung in den ersten Prototypen, was teils aus Zeitgründen in einer Gruppe geschah, und hatten anschließend die Möglichkeit die Anwendung selbst unter Anleitung zu testen. Die Bewertung des ersten Prototypen erfolgte anhand von Fragebögen, die an die Therapeuten ausgeteilt wurden. Geprüft wurde zunächst, ob die umgesetzte Funktionalität den anfänglichen Anforderungen der Therapeuten genügte und wenn nein, wie sie in einem zweiten Prototypen verfeinert werden könnten. Die Logopäden sollten daher entscheiden, welche der fünf umgesetzten Funktionalitäten für einen Einsatz der Anwendung in der Therapie benötigt würden, und diese als wichtig markieren. Funktionen die keinen Mehrwert für die Therapie bringen, sollten als unnötig markiert werden. Tabelle 6.1 verdeutlicht die Anzahl aller Zustimmungen und Ablehnungen zu den Funktionen. Bei Funktionen die weder als wichtig noch als unnötig markiert wurden, zählten die Antworten zu „Keine Meinung“. Es wird deutlich, dass die große Mehrheit der Therapeuten vor allem die Möglichkeit am sinnvollsten fand, einzelne Körperteile in der Anwendung ein- und ausblenden zu können. Ebenso positive Resonanz erhielt die Funktion, vorher gewählte Laute abspielen zu können, zu dem nur drei Therapeuten keine Meinung besaßen. Mehrheitlich war man auch der Auffassung, dass es wichtig sei, Laute über das Mikrofon eingeben zu können. Diese drei Funktionen sollten demnach im zweiten Prototypen grundlegend erhalten bleiben oder gar noch verfeinert werden. Ein interessante Diskrepanz bestand in den Äußerungen zu der Funktion, die Zunge mithilfe von Schieberegler selbst zu bewegen. Zwar waren über die Hälfte der Befragten demgegenüber positiv gestimmt, nur eine Person war allerdings gleichzeitig für eine Zungenbewegung abhängig von ihren Muskeln. Die restlichen Logopäden empfanden es als unnötig, die Muskeln bei der „freien“ Bewegung einzubeziehen oder hatten diesbezüglich keine Meinung.

¹SpeechRecognition Library, <https://pypi.python.org/pypi/SpeechRecognition/> (letzter Aufruf am 09.11.2014)

<i>Fragebogen: Funktionen</i>	<i>Wichtig (Anzahl Antworten)</i>	<i>Unnötig (Anzahl Antworten)</i>	<i>Keine Meinung (Anzahl Antworten)</i>
Vorausgewählte Laute abspielen können	9	0	3
Laute über das Mikrofon eingeben können	7	3	2
Die Zunge mithilfe der Maus selbst bewegen können	8	2	2
Die Muskeln bei einer freien Bewegung mit einbeziehen können	1	5	6
Einzelne Körperteile ein- und ausblenden können	11	0	1

Tabelle 6.1: Fragebogen: Welche vorgestellten Funktionen wären Ihnen praktisch wichtig, auf welche könnten Sie verzichten? Mehrfach-Antworten waren hier möglich

Gefragt wurde weiterhin nach dem Wert der anatomischen Korrektheit einer Zungenstellung gerade im Zusammenhang beim Abspielen von Lauten. Hier waren neun von zwölf Logopäden der Meinung, dass die Bewegung anatomisch exakt vorgeführt sein muss. Zwei waren der Meinung, dass es zu kleinen anatomischen Abweichungen kommen darf, die trotzdem zu einem realistischen Ergebnis führen. Dieses Ergebnis unterstrich die Wichtigkeit, dass im nachfolgenden Prototypen die Lautbildung noch viel deutlicher dargestellt werden musste und dafür die bisherigen Bildmaterialien als Vorlage nicht reichten. Zusätzlich müssten in der Lautliste nicht nur zwischen Lautpaaren ausgewählt werden können, sondern Lenis- und Fortislaute mussten auch unabhängig voneinander abgebildet werden können.

Die nächsten Fragen des Fragebogens dienten der Bewertung, ob und wie hilfreich die Therapeuten den ersten Prototypen für ihre Therapie empfanden und wo die Verbesserungsmöglichkeiten lagen.

Auf einer Skala von 1 (nicht hilfreich) bis 5 (sehr unterstützend) gaben sechs Logopäden an, die Anwendung insgesamt zu diesem Zeitpunkt als unterstützend (Skalenwert 4) für ihre Therapie einzuschätzen. Zwei Logopäden betrachteten sie sogar als sehr unterstützend. Gerade für Patienten mit gutem räumlichen Vorstellungsvermögen biete die Anwendung eine visuelle Unterstützung, sich die Lautbildung vorzustellen. Die Zielbewegungen bzw. Übungen seien dadurch verständlicher. Gerade die 3-D-Perspektive mit uneingeschränkter Sicht auf die Artikulationsorgane sei eine Verbesserung zu bisheri-

gen Bildmaterialien. Motivierend sei die Anwendung auch für Patienten mit technischer Affinität, wie Kinder und Jugendliche, die sowieso nur mit neuester Technik umgehen wollen. Mit einer solchen Anwendung würde eine unterhaltsamere Übungsstunde geboten werden, die weniger anstrengend für Kinder sei.

Trotz der überwiegend sehr guten Resonanz (kein Logopäde fand den Prototypen als nicht hilfreich), wurden auch einige Gründe gegen eine solche Anwendung aufgeführt. Bei kognitiven Einschränkungen würde die Anwendung keine Verständnishilfe sein, eine Therapie am Bett oder bei Hausbesuchen, wie es bei Schlaganfall-Patienten der Fall ist, könne deshalb eine solche Anwendung nicht mit einbeziehen. Es wurden auch Bedenken geäußert, dass gerade jüngere Patienten die Abbildung der Bewegungsabläufe als zu kompliziert empfinden.

Es wird offensichtlich, dass sich viele Aussagen auf grundsätzliche Probleme unabhängig vom ersten Prototypen bezogen, weshalb es nicht zu erwarten war, dass sie diese nach der Umsetzung des finalen Prototypen stark verändern würden oder im Rahmen dieser Arbeit angegangen werden konnten. Für die Verbesserung des ersten Prototypen wurde seitens der Logopäden dennoch zusätzliche Vorschläge in Bezug auf das Modell die Funktionalität angegeben.

Gewünscht wurde sich eine visuell stärkere Differenzierung von hartem und weichem Gaumen bzw. Gaumensegel für das anatomische Modell, da sie ein wichtiger Artikulationsort der palataler und uvularen Laute seien. Auch die Lippen müssten erkennbar gestaltet sein, da labiale bzw. bilabiale Laute nur mithilfe der Lippenstellung korrekt verdeutlicht werden könnten. Die Bewegungen von Gaumensegel und Lippen müssten demnach ebenso korrekt angepasst werden. Weiterhin war es von Interesse die Zungenruhelage in die Animationen mit einzubeziehen. Diese Zungenstellung ist eine wichtige Übung der myofunktionellen Therapie und Voraussetzung einer korrekten Lautbildung. Eine Logopäde gab an, dass die Motivation gerade bei Kindern gesteigert werden könne, wenn die Modelle insgesamt kindgerechter ausgelegt würden. Für den weiteren Prototypen wurde dieser Ansatz jedoch nicht weiterverfolgt, da zum einen die anatomische Korrektheit der Modelle angenommen wurde und diese bei einer kindgerechten Anpassung der Modelle sicherlich verloren gehen würde. Zwar ließen sich die Modelle in Blender mit wenig Aufwand austauschen, dennoch spielt es für die Anwendung keine Rolle, ob es sich bei den Modellen um eine erwachsene oder eine kindliche Anatomie handelt.

Hinsichtlich der Animationen hielten einige Logopäden eine Anpassungsmöglichkeit der Artikulationsgeschwindigkeit als sinnvoll, da die Animationen bei den gegebenen Voreinstellungen teils nur zwei Sekunden abliefen. Zusätzlich wurde vorgeschlagen, nicht nur einen Laut abspielen lassen zu können, sondern auch seine koartikulierten Umgebungslaute bzw. Silben oder einzelne Wörter. Argumentiert wurde damit, dass beim spontanen Sprechen die Laute nicht einzeln auftreten, sondern immer im Zusam-

menhang mit anderen stehen die ihn in seiner Bewegung und Aussprache beeinflussen. So müssten die Unterschiede herausgestellt werden, ob ein Laut initial, medial oder final im Wort auftritt. Weitergehend wurde genannt, dass ergänzend zu der visuellen Verdeutlichung der Laute auch auditive Stimuli angeboten werden könnten, um tatsächlich alle Wahrnehmungskanäle des Patienten mit einzubeziehen.

Bei der Mikrofon-Anbindung erwähnten Therapeuten, dass ein von der Spracherkennung als richtig oder falsch erkannter Laut dem Patienten zusätzliche Rückmeldung geben würde, wie ein Laut richtig zu bilden sei.

Eine gänzlich neue Funktion, die vorgeschlagen wurde, war den physiologischen Schluckvorgang zu visualisieren, der sich von dem des kindlichen deutlich unterscheidet und gerade auch für die Behandlung von Schluckstörungen eine wichtige Übungsgrundlage darstellt.

6.3 Umsetzung des finalen Prototypen

Die Umfragen ergaben zum einen, dass das „freie“ Bewegen der Zunge zwar gewünscht wurde, zum anderen die Verknüpfung der Bewegung mit den Muskeln unnötig war. Diese Funktion musste dementsprechend angeglichen werden. Die Auswahlmöglichkeit bei den Lauten musste auf die der deutschen Sprache angepasst und bei ihrer Animation ein höherer Wert auf die anatomische Korrektheit gelegt werden, was wiederum verbesserte Modelle voraussetzte. Zusätzlich sollten koartikulierte Laute berücksichtigt und der Schluckvorgang visualisiert werden können. Auf Grundlage dieser Auswertung und den genannten Verbesserungsmöglichkeiten konnten die grafische Oberfläche und alle gegebene Funktionalitäten des ersten Prototypen erweitert oder ergänzt werden.

Wie in Abb. 6.7 zu sehen, wurde die grundlegende Einteilung des Interfaces in drei Regionen im finalen Prototypen übernommen und um eine weitere Region ergänzt (4). Das 3D-View-Fenster ist noch immer interaktives Element der Anwendung, das Gesamt-Modell musste jedoch angepasst werden. Da nun der Schluckvorgang animiert wurde, musste das Modell um alle daran beteiligte pharyngo-laryngeale Organe ergänzt werden. Das wichtigste Organ, die Epiglottis, hebt sich dabei mit einer anderen Materialfarbe von den anderen ab. Die drei extrinsischen Muskeln der Zunge, die die „freie“ Bewegung der Zunge im ersten Prototypen bestimmten, wurden aus dem neueren Kopfmodell herausgenommen, da sie mit keiner Funktion mehr verknüpft waren. Das vorher eher rudimentäre Modell der Mundhöhle wurde im finalen Prototypen ersetzt durch ein Modell des Hartgaumens mit Velum (Gaumensegel), was eine eigene Animation für jeden Laut erhielt. So konnte je nach Laut die Abgrenzung von Mund- und Nasenhöhle und damit einhergehend auch der Verlauf der Ausatmung bei der Artikulation verdeutlicht werden.

Für eine exaktere Animation der Zunge, was als Anforderung aus der Umfrage hervorging, benötigte das Zungenmodell symmetrische Eigenschaften und ein regelmäßi-

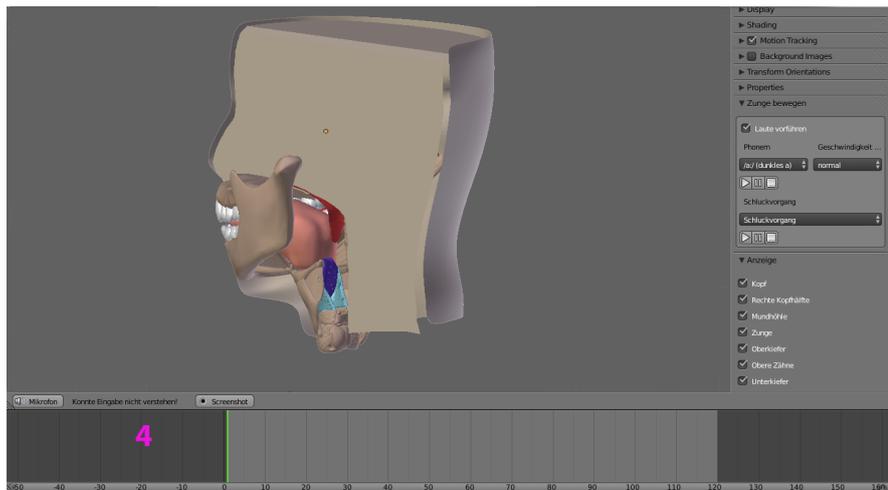


Abbildung 6.7: Bedienoberfläche und Modell des finalen Prototypen. Ansicht auf das Gesamtmodell

ges Polygonnetz, das feinere Transformationen einzelner Bereiche erlaubte. Es mussten klare Grenzen zwischen Zungenspitze, -körper und -grund gezogen werden und die Seitenränder symmetrisch transformiert werden können. Deshalb wurde ihr Modell mithilfe eines SubdivisionSurface-Modifiers verfeinert und quadranguliert (Abb. 6.8b). Auf dem Zungenmodell konnten damit Vertex-Gruppen für Zungenspitze, -Körper und -Grund erstellt werden, die zusätzlich zum ebenso verwendeten MeshDeform-Modifiers eine feiner abgestufte Transformation der Zunge erlaubten. Dadurch konnten für eine Animation der Zunge Bereiche auf der Zunge nicht nur mithilfe des Modifiers transformiert werden, sondern gleichzeitig mithilfe der Zunge selbst, was insgesamt in einem Shape Key für die Zunge festgehalten wurde. Abbildung 6.9a und 6.9b verdeutlichen diesen Zusammenhang anhand zwei ausgewählter Vertex-Gruppen für die Transformation der Zungenspitze. Das dem ersten Prototypen basierenden Kopfmodell wurde an seiner vertikalen Symmetrieachse so vervollständigt, dass die rechte Kopfhälfte aber jederzeit ein- und ausgeblendet werden konnte. Eine Abeckung für die obere Hälfte des Kopfes sollte den darunter liegenden Hohlraum verbergen und zugleich den Fokus auf die Visualisierung der Artikulationsorgane legen. Auch die Lippen wurden so modelliert, dass sie nicht nur visuell stärker hervortraten, sondern auch differenzierter bilabiale Laute wie [p] und [b] formen konnte. Gerade zur Vorführung des Lautes [f] werden erkennbar vorgestülpte und deutlich gerundete Lippen benötigt. Die Visualisierung des Phonationsstroms, d. h. Auatemluft beim Ausführen der Laute, wurde mithilfe eines Partikelsystems zwischen zwei Modellen und einem Kontrollmodell realisiert. Diese Modelle sind für den Benutzer nicht sichtbar, sondern nur der resultierende Partikelstrom im 3D-View-Fenster. Die Modellierung von Nasenlöchern im Kopfmodell war deshalb Voraussetzung für die Verdeutlichung des Phonationsstroms bei nasalen Lauten.

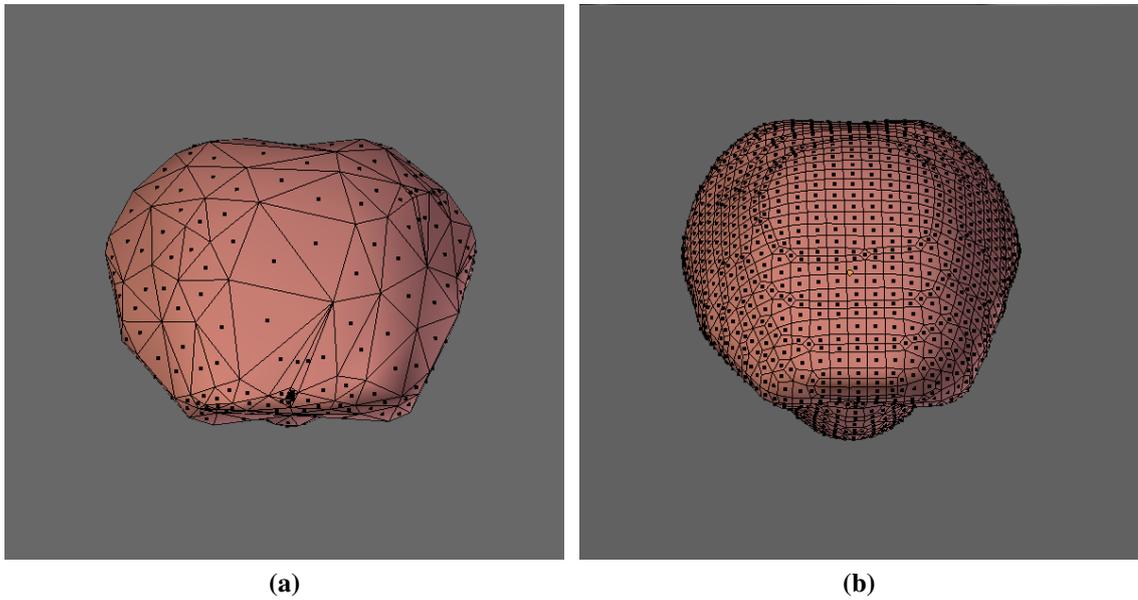


Abbildung 6.8: Die beiden Polygonnetze des Zungenmodells vom ersten und finalen Prototypen im Vergleich. (a) erster Prototyp, (b) finaler Prototyp

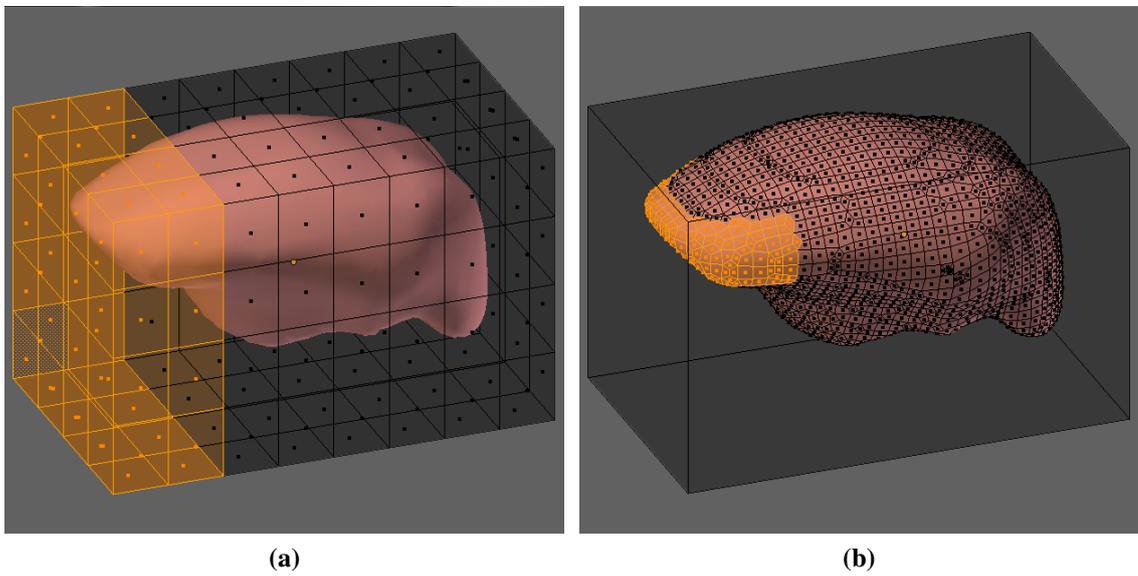


Abbildung 6.9: Zwei Vertex-Gruppen dienen dem verfeinerten Transformieren der Zungenspitze. (a) Vertex-Gruppe auf dem Polygonnetz des MeshDeform-Modifiers, (b) Vertex-Gruppe auf dem Polygonnetz der Zunge

Das rechte UI-Panel bietet wie im ersten Prototypen die entsprechenden Bedienelemente um Laute oder den Schluckvorgang vorführen zu lassen, die Zunge frei zu bewegen oder Teile des Modells ein- und ausblenden zu lassen. Der Header des 3D-View-Fensters enthält wieder die Mikrofon-Anbindung, wurde aber um ein weiteres Element ergänzt, das es erlaubt ein Bild der aktuellen Modell-Ansicht für ein hochwertiges Screenshot zu rendern. Diese Umsetzung dieser Funktion folgte nicht direkt aus der Zwischenevaluation des ersten Prototypen, sondern war vielmehr ein Teil der technischen Anforderung *für* den ersten Prototypen.

Die neu hinzugekommene Region (4) entspricht dem Timeline-Fenster von Blender, mit der Möglichkeit die Bewegungen der Objekte, sprich die Animation, über die Zeit zu kontrollieren. So kann per Maus jede pausierte Animation vor- und zurückgespult werden.

6.3.1 Zunge bewegen

Im finalen Prototypen kann der Therapeut Bewegungen der Zunge auf zwei verschiedene Arten ausführen. Er kann aus einer Liste einen abzuspielenden Phon auswählen, zusammen mit einer Geschwindigkeit oder einer Tonausgabe, oder die Zunge mit Schieberegler manuell in verschiedene Raumrichtungen bewegen.

Auch wenn die Liste mit vorzuführenden Lauten insgesamt nur eine Auswahl darstellt, da sie einige Laute nicht enthält wie [h] oder Diphthonge, wurde sie weitestgehend mit den übrigen im Deutschen vorkommenden Lauten ergänzt. Im Gegensatz zu der vorgegebenen Liste des ersten Prototypen wird bei dieser Liste nicht nur zwischen Fortis- und Lenislauten unterschieden, sondern oft auch zwischen zwei möglichen Arten der korrekten Lautbildung. In Abbildung 6.10 werden diese Laute präsentiert, beispielsweise das Lautpaar [p] und [b], die nun als Fortis- und Lenislaut einzeln ausgewählt werden können. Der Laut [d] kann sowohl apikal (mit der Zungenspitze) und dorsal (mit dem Zungenrücken) gebildet werden, so wie der Laut [r] mit der Zungenspitze oder mit dem Gaumenzäpfchen ([ʀ]) gebildet werden kann. Zusätzlich lässt sich aus der Liste auch zum Vorspielen die Zungenruhelage auswählen.

Eine weitere Liste im Panel lässt den Therapeuten die Geschwindigkeit der Animation ändern (Abb. 6.11), die wie im letzten Prototypen standardmäßig eine Vorschau von 60 Frames beträgt. Sie kann nun auf die Hälfte verlangsamt (120 Frames) oder doppelt so schnell (30 Frames) abgespielt werden. Bei normaler Geschwindigkeit ist zusätzlich der Phonationstrom zu beobachten. Wenn das Listenelement gewählt wurde, die Animation in normaler Geschwindigkeit zusammen mit Ton abzuspielen, werden in einer neuen Liste zwei Möglichkeiten für den Tonausgabe angegeben (Abb. 6.12a). Wenn der gewählte Laut ein Konsonant ist, lässt er sich dadurch beispielhaft in Verbindung mit dem Vokal [a] abspielen in Form von *a_Laut_a* oder *Laut_a*.

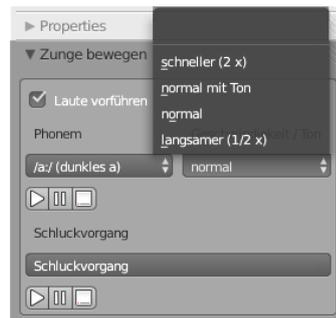


Abbildung 6.11: Liste auszuwählender Geschwindigkeiten

zur Zungenstellung in schnellen Bewegungen animiert werden müssen. Die Werte der einzelnen Keyframes der Gruppen wurden bei der Umsetzung manuell getestet und anschließend im Python-Skript vordefiniert. Die Shape Keys der restlichen Laute werden wie im vorangegangenen Prototypen automatisch gesetzt.

Als Vorlage mit Abbildung und Beschreibung der Weise, wie ein Laut gebildet wird, diente hier Fiukowski [Fiu10]. Darin wird genau auf alle beteiligten Artikulationsorgane und ihre Eigenschaften eingegangen.

Zungenruhelage

Die Zungenruhelage beschreibt die Stellung der Zunge, wenn der Mund entspannt geschlossen ist, d. h. die Zahnreihen sich in der Ruheschwebe befinden, und durch die Nase geatmet wird. Sie ist Voraussetzung für ein korrektes Sprechen und Schlucken, weshalb sie als Teil der myofunktionellen Therapie häufig geübt wird. Das Animationsergebnis ist in Abbildung 6.12b zu betrachten. Das vordere bis mittlere Drittel des Zungenkörpers liegen am Hartgaumen an, die Zungenspitze am Alveolardamm der oberen Zähne.

Tonausgabe

Die Tonausgabe ist eine Möglichkeit, die beispielhaft nur bei normaler Animationsgeschwindigkeit und mit dem Vokal [a] geboten wird, da nur Audiodateien mit diesem Vokal vorlagen. Sie stammen von der Wikipedia-Übersicht der IPA-Zeichen² und wurden manuell nachbearbeitet. Im Sinne der Koartikulation verdeutlichen die Audiobeispiele die Anbahnung eines Lauts in Abhängigkeit von seiner medialen (*a_Laut_a*) und initialen (*Laut_a*) Stellung in einer Silbe. Die Dateien wurden auch in dieser Form abgespeichert, sodass sie beim Abspielen des Lautes als neue Audio-Sequenz in Blender angelegt werden und die Animationskurve der Lautbewegung an die der Audioausgabe

²Wikipedia, http://de.wikipedia.org/wiki/Liste_der_IPA-Zeichen, (letzter Aufruf 09.11.2014), User: Denelson83

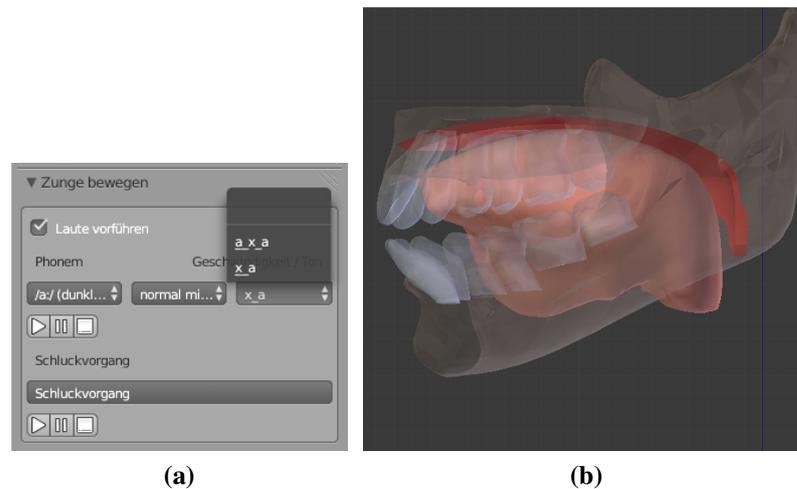


Abbildung 6.12: (a) Liste auszuwählender Tonausgaben. In der Liste steht das x für den ausgewählten Laut und das a für den koartikulierten Laut, (b) Stellung der Zunge während der Zungenruhelage. Der Kiefer ist in dieser Ansicht transparent

angehängt werden konnte. Die Modellbewegungen erfolgten damit in Abhängigkeit von der Audiokurve und synchron mit der Tonausgabe.

Phonationsstrom

Die Visualisierung der Ausatmung bei der Lautbildung ist nicht direkt den Anforderungen an den zweiten Prototypen entnommen, soll aber ergänzend verdeutlichen, wie intensiv ein Laut gebildet werden muss.

Der Phonationsstrom wird bei gewählter Normal-Geschwindigkeit visualisiert und bildet die Intensität des gesprochenen Lauts ab (Abb. 6.13). Dafür wurde ein Partikelsystem verwendet mit einem Start- und Zielobjekt, zwischen denen sich die Partikel bewegen, und einem Kontrollobjekt, das die Eigenschaften des Partikelverhaltens bestimmt. Die Partikel der drei Arten von Expirationsluft unterscheiden sich in der Anzahl und ihrer Lebenszeit, d. h. wie lange sie visualisiert werden, weiterhin lassen sich Start- und Endframe der Partikel setzen. Diese Werte mussten zunächst aufeinander abgestimmt und manuell angepasst werden. Zusätzlich wurden die Partikel einer Art noch farblich codiert, um eine klare Differenzierung zu erreichen. Ein konstanter Phonationsstrom weist grüne Partikel auf, die sich langsam und vermehrt Richtung Ziel bewegen, aber eine breite Fläche einnehmen. Ein sanfter Phonationsstrom enthält wenige gelbe Partikel, die nur keine hohe Lebenszeit besitzen und deshalb nur kurz erscheinen; ein kräftiger besteht aus vielen orangenen Partikeln gehäuft auf kleiner Fläche, die sich schnell, aber mit nur kurzer Lebenszeit zum Ziel bewegen.

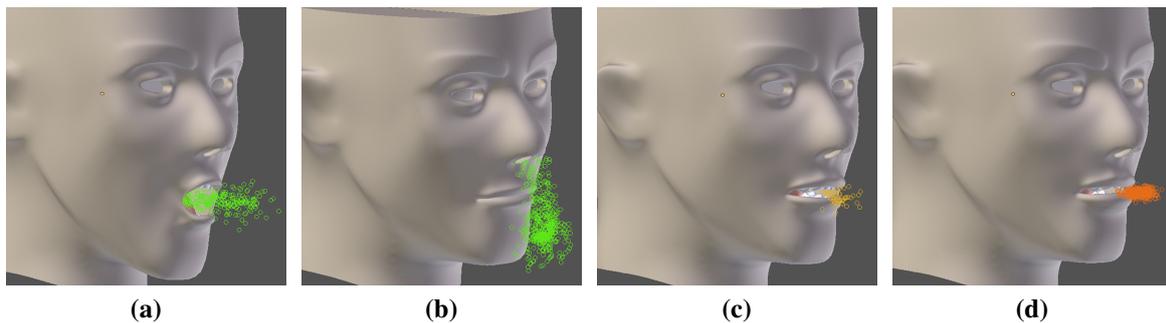


Abbildung 6.13: Konstanter, konstant nasaler, leichter und kräftiger Phonetationsstrom im Vergleich. (a) konstant beim Laut [o], (b) nasal beim Laut [m], (c) sanft beim Laut [b], (d) kräftig beim Laut [t]

„Freie“ Bewegung

Sobald das Kontrollkästchen für das Vorspielen von Lauten deaktiviert ist, erscheint das Menü um die Zunge mithilfe von Schieberegler zu bewegen (Abb. 6.14a) und in eine bestimmte Stellung zu bringen. Das Prinzip folgt dem des ersten Prototypen, nur dass hier die Bewegungen nicht über die Visualisierung der Muskeln erfolgt: Die Regler greifen auf bestimmte Shape Keys der Zunge zu, die vorher mithilfe eines MeshDeform-Modifiers, mit Ursprung im Zungengrund, gesetzt wurden.

Der erste Schieberegler lässt den Mund und alle daran beteiligten Modelle öffnen, die Zunge befindet sich dann an den unteren Zahnflächen. Über zwei neu hinzugekommene Menüs können Zungenkörper und Zungenspitze separat transformiert werden, nach oben, unten, links oder rechts. Der Zungenkörper kann zusätzlich noch entlang der lokalen X-Achse eingezogen oder rausgeschoben sowie breit und schmal gemacht werden (Skalierung entlang der lokalen Z-Achse). Durch die Verknüpfung mehrerer Bewegungen können so mithilfe des Modells auch Übungen der myofunktionellen Therapie veranschaulicht werden.

Damit die Zunge während der Transformation nicht an die oberen oder unteren Frontzähne stoßen kann, wird für jede Verschiebung des Reglers eine einfache Art der Kollisionserkennung ausgeführt. Sie findet zwischen den Eckpunkten der vorderen Zunge und den Flächen der Kollisionsmodelle statt. Dafür wurden zwei Kollisionsobjekte benötigt, die eine einfache Geometrie besitzen und sich an die Zähne anschmiegen. Sie sind in Abb. 6.15b veranschaulicht. Das Kollisionsobjekt der unteren Zähne bewegt sich bei Verschiebung des Kiefers mit. Für einen einfachen Zugriff werden die Flächenelemente der Kollisionsmodelle als Tupel in Form von

$$face_{bb} = (fIndex, fNormal, [fVertex_1, fVertex_2, \dots])$$

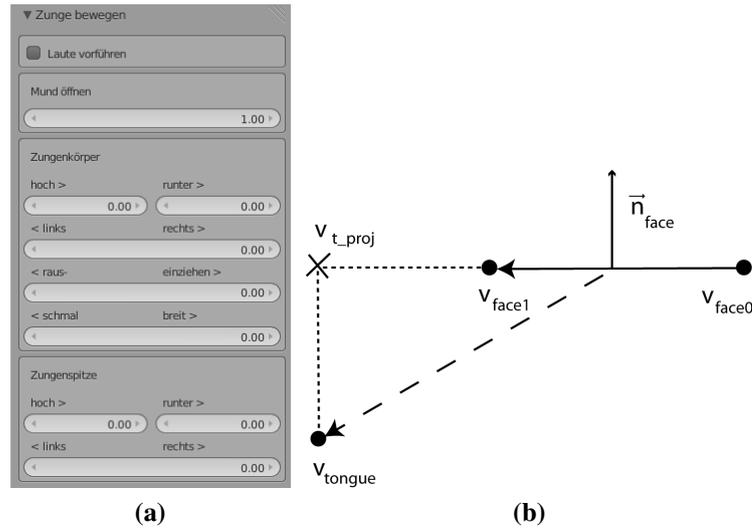


Abbildung 6.14: (a) Bedienelemente um die Zunge frei zu bewegen, (b) Skizze des Kollisions-tests von der Seite. Mithilfe des Skalarprodukts wird der Vertex der Zunge v_{tongue} auf die Ebene des Flächenelements projiziert

in einer Liste abgelegt, wobei die Flächennormale $fNormal$ und Flächeneckpunkte $fVertex$ 3D-Vektoren in Weltkoordinaten darstellen, und $fIndex$ der Index der Fläche im gesamten Polygonmodell ist. Die Eckpunkte des Polygonnetzes der Zungenspitze werden ebenfalls in einer Liste verwaltet, sodass jeder dieser Eckpunkte \vec{v}_{tongue} mit jeder Fläche $face_{bb}$ der vereinfachten Zahnmodelle auf Kollision getestet werden kann. Dazu wird zunächst mithilfe des Skalarprodukts und der Flächennormale \vec{n}_{face} überprüft, ob sich \vec{v}_{tongue} im vorderen oder hinteren Halbraum der Fläche $face_{bb}$ befindet:

$$dotProduct = ((\vec{v}_{tongue} - \vec{v}_{face0}) \circ \vec{n}_{face})$$

Falls $dotProduct > 0$ ist, befindet sich die Zunge im vorderen Halbraum der Fläche. Falls $dotProduct \leq 0$ ist, muss eine Fallunterscheidung getroffen werden, da nicht zwangsläufig eine Kollision stattfindet. Dafür wird \vec{v}_{tongue} mithilfe des Skalarprodukts auf die Ebene des Flächenelements projiziert:

$$dotP = (\vec{v}_{face1} - \vec{v}_{face0}) \circ (\vec{v}_{tongue} - \vec{v}_{face0})$$

Falls nun $dotP \geq 0$ ist, ergibt sich der projizierte Punkt aus

$$\vec{v}_{tongue_{proj}} = (dotP \cdot (\vec{v}_{face1} - \vec{v}_{face0})) + \vec{v}_{face0}$$

Mit $|\overrightarrow{v_{face0}v_{tongue_{proj}}}| \leq |\overrightarrow{v_{face0}v_{face1}}|$ liegt der Punkt in der Fläche. Ansonsten muss die Rechnung für einen den zweiten Punkt der Fläche \vec{v}_{face2} wiederholt werden. Sobald eine Kollision stattgefunden hat, werden die Slider auf ihren ursprünglichen Wert zurückgesetzt.

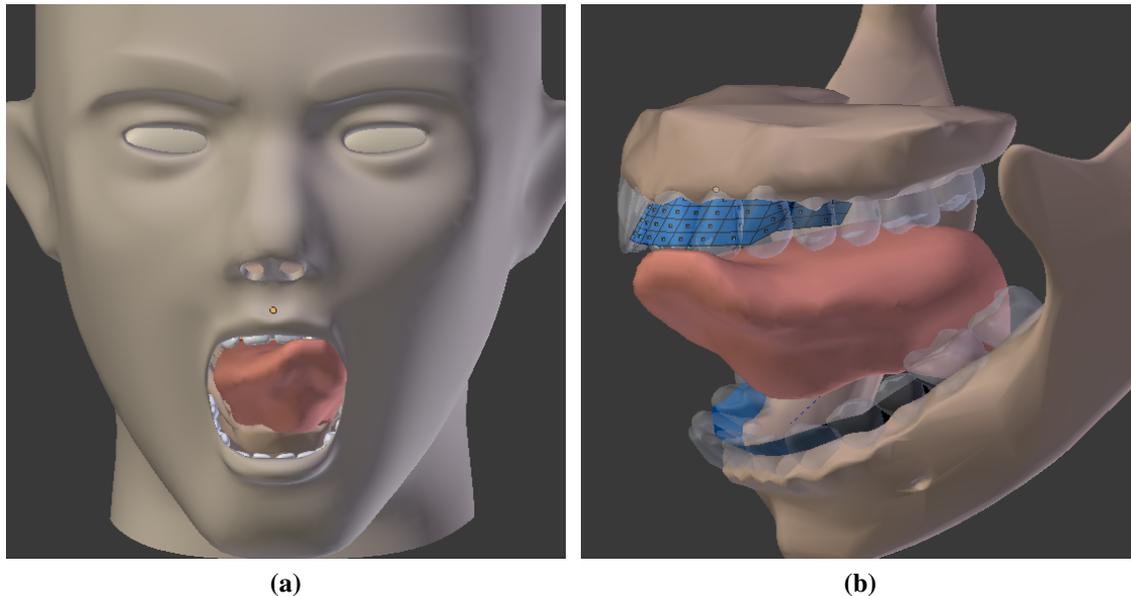


Abbildung 6.15: Die Kollisionserkennung findet während der Bewegung statt. (a) Die Zunge wird nach oben und links verschoben, die Zungenspitze wird nach oben an die Zähne bewegt, (b) Kollisionsmodelle der oberen und unteren Zähne (blau)

6.3.2 Schluckvorgang

Im gleichen Panel wie für das Vorführen von Lauten befinden sich eigene Bedienelemente für die Simulation des Schluckvorgangs. Dieser musste mit Vorlage der Beschreibungen von Schindelmeiser [Sch10] nach-animiert werden. Jedes beteiligte Modell der Animation erhielt daher fünf Shape Keys, die die fünf Phasen des Schluckens repräsentieren. Abbildung 6.16 verdeutlicht reduziert vier der fünf Phasen. Der Bolus bzw. Speisebrei wird hier dargestellt durch ein Objekt mit Partikelsystem, dessen Positionen und Transformationen für die Animation bereits vordefiniert wurden. Die Bewegungen der anderen Modelle werden über Shape Keys zu vorher festgelegten Frames gesteuert. Die Epiglottis ist farblich hervorgehoben.

6.3.3 Mikrofon-Anbindung

Die Funktionalität des Mikrofons ist im Header des 3D-View-Fensters zu finden. Bei Drücken des Buttons kann wie im letzten Prototypen ein Wort eingesprochen werden. Sobald dieses erkannt wird, werden alle Buchstaben des Wortes in einer Auswahlliste angezeigt, die sich einzeln abspielen lassen. Ein neues Kontrollkästchen bietet aber auch die Funktion, alle Buchstaben des Wortes nacheinander abspielen zu lassen. Jeder Laut umfasst dann 15 Frames, das Vorschau-Ende ist demnach abhängig von der Länge des

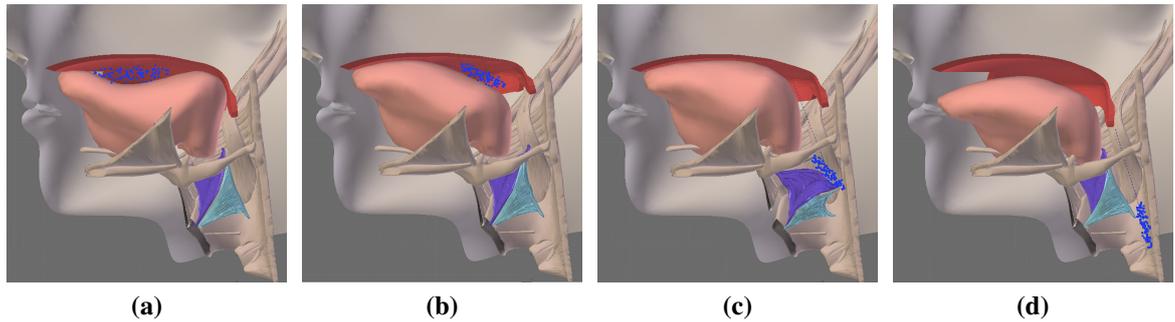


Abbildung 6.16: Die Schluckbewegungen werden in vier Phasen eingeteilt. Der Bolus ist dunkelblau dargestellt, die Epiglottis ist lila dargestellt. (a) Vorbereitungsphase: Der Bolus liegt in der Zungenschüssel, (b) Orale Phase: Das Velum ist gehoben, (c) Schluckreflex und pharyngeale Phase, (d) Ösophageale Phase

Wortes. Die Animation besteht dabei aus der Hintereinander-Ausführung der einzelnen Laute.

6.3.4 Render-Funktion

Der ebenfalls im 3D-View-Header angelegte Screenshot-Button, erlaubt es per Knopfdruck in die Cycles-Renderengine zu wechseln, bei der ein qualitativ hochwertiges Bild gerendert wird. Eine eingestellte oder pausierte Zungenstellung aus der Animation kann damit zum Ausdrucken vorbereitet werden (Ab. 6.17). Bei nochmaligen Drücken des Buttons wird wieder in den normalen Modus gewechselt.



Abbildung 6.17: Modellansicht in der Cycles-Renderengine von Blender mit 20 Samples

Kapitel 7

Evaluation und Bewertung

Eine Evaluation des finalen Prototypen fand über ein Experteninterview mit einer Logopädin statt, die im klinischen Bereich tätig ist. Vorderranges Ziel der Befragung war es herauszufinden, ob die Umsetzung des finalen Prototypen die Aussagen der Zwischenevaluation unterstützt. Weitergehend stellte sich die Frage, wie der Einsatz dieser Anwendung in der Therapie aussehen könnte. Im anschließenden Abschnitt werden die gegebenen Argumente bewertet und die Anwendung einem Vergleich mit den konzeptähnlichen Verfahren *Baldi* und *ARTUR* unterzogen.

7.1 Evaluation

Die Logopädin wurde zunächst durch die Anwendung geführt, um alle verfügbare Funktionalität kurz zu verdeutlichen. Diese wurden dann aus logopädischer Sicht im Verlauf des Interviews hinsichtlich des Mehrwerts gegenüber den traditionellen Therapiemethoden bewertet.

Aus dem Gespräch ging hervor, dass die Umsetzung der Lautvorführung über Animationen im Vergleich zu einfachen Bildkarten zum einen dynamisch, und zum anderen deutlich besser visualisiert sei, da die Bildkarten meistens vereinfachte schwarz-weiß-Zeichnungen sind. Übungen können zwar auch dynamisch an sich selbst vorgeführt werden, bieten aber keine freie Sicht auf die wichtigen Strukturen und den Bewegungsablauf. Dieses Problem wird in der Anwendung durch die Möglichkeit behoben, indem in der 3D-Visualisierung Modelle ein- und ausgeblendet werden können.

Die Audio-Ausgaben unterstützen das Therapievorgehen, Laute immer zusammen in der Silbe mit einem Vokal zu üben. Gerade im neurologischen Bereich werde die Sprache bzw. das Sprechen über Silbentraining stimuliert, dessen Behandlungsziel das Sprechen von Wörtern ist. Dabei sei die Tonausgabe der Anwendung in Form von *Laut_a* jedoch sinnvoller als *a_Laut_a*. Erstere Form übt genau das Anbahnen des Lautes, was

Voraussetzung für seine korrekte Artikulation ist. Die visualisierten Phonationsströme stellten sich aus dem logopädischen Blickwinkel als gute Möglichkeit heraus, die verschiedenen Artikulationsintensitäten zu verdeutlichen. Feinheiten würden so herausgestellt werden, die sich nicht an der Lippenstellung unterscheiden würden wie beim Laut [b] und [p], und auch Nasallaute würden so sehr gut beschrieben.

Die Aufnahmen von Wörtern über das Mikrofon wurde eher kritisch von der Logopädin beurteilt. Auf der einen Seite müsse geklärt werden, ob die Spracherkennung zwischen potentiell falsch und richtig ausgesprochenen Wörtern unterscheiden könne. Das Programm könne sonst vermutlich undeutlich gesprochene Worte erkennen, die aus Sicht des Logopäden nicht deutlich genug ausgesprochen wurden, oder Worte mit vertauschten Buchstaben akzeptieren. Andersherum könnten korrekt ausgesprochene Worte gar nicht erkannt werden. Auf der anderen Seite unterstütze die Funktion die Therapie auf Wortebene und orientiere sich dadurch an den alltäglichen Übungen. Die Möglichkeit, die Eingabe zusätzlich zu visualisieren, wäre dann ein vorteilhafter Nebeneffekt. Diesbezüglich sei auch die im finalen Prototypen eingeführte Zeitleiste ein sinnvolles Mittel, falls die Bewegungen in Echtzeit zu schnell gingen, da damit sich zeitlich durch ein Wort oder einen Laut bewegt werden kann.

Differenzierte Äußerungen wurden auch gegenüber dem Bewegen der Zunge mithilfe von Schieberegler getroffen. Die Funktion sei in dem Sinne überflüssig, dass der Logopäde die damit vorgesehenen Bewegungen der myofunktionellen Therapie auch an seinem eigenen Mundbild vorführen kann. Die dahinterliegenden Strukturen sind nicht von Bedeutung, wenn die Zunge nur in die Mundwinkel bewegt werden oder kreisrunde Bewegungen auf den Lippen ausführen soll. Daher bringe auch die zusätzliche Visualisierung keinen Mehrwert.

Was den Einsatz der Anwendung im Allgemeinen betrifft, so könnte sich das Verwenden der „neuesten Technik“ motivierend auf Kinder und Jugendliche auswirken, mit dem positiven Effekt, dass die Anwendung selbst ausprobiert werden und für ein Üben zu Hause genutzt werden könnte. Die Anwendung wäre demnach insofern unterstützend, als dass sie auch unabhängig vom Logopäden funktionieren könnte. Während der Therapiesitzung könne die Anwendung als ergänzendes Material eingesetzt werden, ohne dass sie die Therapie ersetzt. Da es in jeder Therapiestunde andere Schwerpunkte gebe, könne das Programm jedoch nur als Teil der Stunde betrachtet werden. Falls es dann eingesetzt wird, müsse es parallel zur Behandlung laufen, der Zweck sei aber nicht, Übungen auf dem Computer mit den üblichen zeitbegrenzt abzuwechseln. Hinzu kommt, dass man auch als Logopäde selbst für eine Computer-Anwendung im Allgemeinen motiviert sein müsse. Oft seien die üblichen Therapiematerialien schneller herauszuholen und einfacher zu handhaben, da sie an Bekanntes anknüpfen und die Erfahrung im Umgang bereits gegeben ist. Aus Sicht der befragten Logopädin müsste daher die Angst der An-

wender berücksichtigt werden, schnell mit technischen Problemen konfrontiert sein zu können.

Der Einsatz der Anwendung hänge gerade auch von dem Schweregrad der Beeinträchtigung der Patienten ab. Der Patient müsse für die Anwendbarkeit der Anwendung nicht nur in das Therapiezimmer kommen können, sondern auch die nötige visuelle Wahrnehmung vorweisen. Beides sei bei bettlägerigen Patienten mit neurologischen Störungen nicht der Fall. Da aus Sicht der Logopädin die Anwendung vor allem die mundmotorischen Übungen unterstützt, sei sie interessant für Patienten mit Dysarthrien, Sprechapraxien und allgemein bei Patienten mit myofunktionellen Störungen und Artikulationsstörungen. Bei der Behandlung von pathologischen Schluckvorgängen könne die Anwendung als Hilfe für vorbereitende Zungenkräftigungsübungen dienen. Dabei wird beispielsweise der Laut [k] eingenommen, aber nicht gesprochen, oder der Laut [t] mit der Zungenspitze wiederholt geübt. Die Simulation des Schluckvorgangs in der Anwendung hebe anatomisch korrekt die wellenförmige Bewegungen der Zunge hervor, da die Zungenspitze erst an den harten Gaumen und der Zungenrücken anschließend an den weichen Gaumen gedrückt wird. Gerade dies könne bei einem Vorführen am eigenen Mundbild nicht eingesehen werden.

7.2 Bewertung

Wie erwartet waren die Äußerungen, die den allgemeinen Einsatz von digitalen Medien betrafen, gleichbleibend zu den Angaben der Zwischenbefragung nach dem ersten Prototypen. Schon bei der ersten Umfrage wurden allgemeine Zweifel am eigenen Umgang mit der Technik deutlich, obwohl nur eine Person seine Computer-Kenntnisse als schlecht einschätzte. Zu beobachten waren dementsprechend positive Äußerungen gegenüber den Einsatz von Technik von eher den jüngeren Logopäden. Vermutlich fehlte hier schlicht die nötige Erfahrung, vor allem da alle Logopäden angaben, keine Art von digitaler Anwendung in ihrer Therapie zu verwenden. Mehrmals bezogen sich die Argumente auch auf die Wirtschaftlichkeit der Anwendung und die technischen Ausstattungen in den Einrichtungen. Da diese Arbeit jedoch zunächst nur von einer prototypischen Anwendung ausgeht, wurden keine ökonomische und hardwareseitige Anforderungen an die Anwendung gestellt.

Die Interaktionsmöglichkeit Laute vorführen zu lassen und gleichzeitig Teile des Modells ausblenden zu können um die Artikulationsorgane aus jedem Blickwinkel einzusehen, bestätigte sich als Vorteil gegenüber einer Verdeutlichung am eigenen Mundbild und der sonst verwendeten Bildmaterialien. Kröger et al. [KBHM10] stellen ein ausgeprägtes Modell der Artikulationsorgane vor, bieten aber nur eine sagittale Sicht auf dieses. Das Konzept von Massaro [mas] und Engwall et al. [EBOK06] präsentieren ein dreidimensionales Artikulationsmodell, was nicht alle Artikulationsorgane vollständig

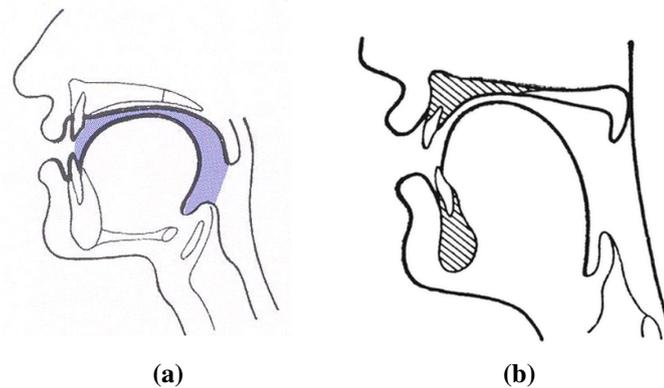


Abbildung 7.1: Sagittalschnitt des Mundraums bei Bildung des Vokals [i]. (a) nach [Sch10, S. 138], (b) nach [Fiu10, S. 374]

abbildet, da die Einzelmodelle anatomisch ungenau visualisiert werden. Dennoch kann die Anforderung an die anatomische Korrektheit in dieser 3D-Visualisierung nur bedingt nachgeprüft werden. Zum einen sind anatomische Strukturen immer individuell verschieden, was sich – wenn auch in nur in minimalen Unterschieden – auf die Parameter der Zungenstellung oder Kieferöffnung auswirkt. Die Laute werden dann auf verschiedene Arten korrekt artikuliert, dementsprechend gibt es nicht nur *eine* korrekte Bildung des Lauts. Zum anderen sind die Bildvorlagen, die man für die Umsetzung der Zielpositionen im Dreidimensionalen verwenden muss, sehr unterschiedlich verdeutlicht. Ein Beispiel ist in Abb. 7.1a und 7.1b gegeben. Im Vergleich wird ersichtlich, dass in Teilbild (a) das Velum nicht gehoben ist bzw. nicht an der Rachenwand anliegt wie in (b). Auch ist der vordere Zungenrücken in (b) deutlich zum vorderen Hartgaumen aufgerichtet und stark gewölbt. Zusammenfassend kann die Visualisierung also immer nur das Prinzip der Lautbildung verdeutlichen. Nach Meinung der Expertin ist die Visualisierung in der Anwendung aber zumindest in diesem Sinne exakt. Durch ein erweitertes Kopfmodell im finalen Prototypen konnten im Gegensatz zu dem des ersten viel deutlichere Lippenstellungen erreicht werden (Abb. 7.2). Gerade bei Übungen des [f]-Lauts wird die richtige Lippenstellung benötigt.

Die Verdeutlichung des Phonationsstroms während der Lautbildung wurde im Experteninterview als neue Funktionalität vorgestellt und positiv beurteilt, obwohl sie keine direkte Anforderung für den zweiten Prototypen darstellte. Für die Umsetzung konnten Mittel genutzt werden, die Blender ohnehin bot (Partikelsystem). Andere Funktionalitäten wie die Anpassungsmöglichkeit der Geschwindigkeit und die Zeitleiste folgten als Anforderung aus dem ersten Prototypen und wurden nicht näher beurteilt, sondern als vorteilhaftes Nebenprodukt hingenommen. Dementsprechend lässt sich daraus folgern, dass sie doch Voraussetzung für eine solche Anwendung sein müssen, damit der Logopäde die Animationen selbst kontrollieren kann.



Abbildung 7.2: Die Lippen sind im finalen Prototypen prägnanter modelliert, sodass Phone wie [p] (a) und [f] (b) deutlicher dargestellt werden können

Was ebenso im Experteninterview positiv aufgenommen wurde, war die Tonausgabe. Sie spricht die auditive Wahrnehmung des Patienten an, wenn ein Laut über eine Silbe angebahnt wird. Es muss erörtert werden, ob die Funktion von Nutzen für den Logopäden ist, wenn er den Laut auch selbst vorsprechen kann. Eine Unterstützung für das Silbentraining würde dann nur durch die gleichzeitige Visualisierung der Bewegungen geboten werden, da hier wieder die Vorteile der 3D-Visualisierung greifen. Hier jedoch stellte sich als Problem heraus, dass ein einfaches Anbinden einer Audiodateien an die Interpolationskurve der Animation in Blender nicht genügte. Die Synchronisation von Ton und Visualisierung führte dazu, dass die Zielstellungen der Artikulationsorgane nur schwach bis gar nicht eingenommen wurden. Zusätzlich müsste die Tonausgabe in Verbindung mit anderen Vokalen möglich sein, die aktuelle Funktionalität beinhaltet lediglich Silben mit dem Vokal [a].

Die Zunge „frei“ bewegen zu können war eine Anforderung, die aus der Beurteilung des ersten Prototypen folgte, jedoch sollte die Bewegung von den Muskeln losgelöst sein. Das Experteninterview bestätigte nicht, dass eine solche Funktionalität tatsächlich von Mehrwert sei. Die Äußerung war, dass gerade myofunktionelle Übungen am Logopäden selbst vorgeführt werden können, weil für diese keine 3D-Visualisierung des Vokaltrakts notwendig sei. In der Hospitation konnte beobachtet werden, dass der Patient bei Ausführung der Übung selbst im Spiegel betrachten muss und nicht unbedingt das Vorführen des Logopäden benötigt. Der Logopäde muss lediglich die Übung anfangs vormachen und beobachtet dann die Bewegungen des Patienten, was jenes Argument untermauert. Der Nutzen der Funktion für die Anwendung wäre demnach nicht gegeben. Weiterhin lässt es der aktuelle Stand des finalen Prototypen noch nicht zu, Bewegungen der Zunge wie in Abbildung 7.3 durchzuführen. Zum einen werden bisher nur Kollisionen zwischen der Zunge und den Kollisionsmodellen der unteren und oberen Frontzähne behandelt und nicht mit der Mundhöhle oder den Lippen. Zum anderen ist die Bedienung über mehrere Schieberegler nicht für solche Transformationen ausgelegt, da sie in diesem Zusammenhang nicht intuitiv erscheint. Zu fragen bleibt

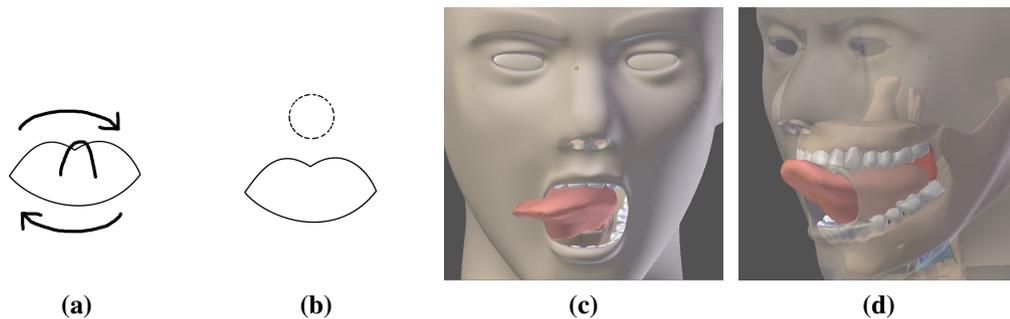


Abbildung 7.3: Myofunktionelle Übungen, die sich mit Schiebereglern kaum realisieren lassen. (a) Kreisen der Zunge an den Lippen, (b) Legen der Zunge zwischen die oberen Frontzähne und Oberlippe, (c) Die Lippen und Mundhöhle werden nicht auf Kollision getestet, (d) Die Zunge lässt sich an die obere Lippe anlegen, da sie mit den oberen Zähnen bereits kollidiert.

deshalb, ob nicht auch solche Bewegungen für die Therapie reichen würden, die nicht als myofunktionelle Übungen ausgelegt sind.

Berechtigte Zweifel betreffen das Aufnehmen von Lauten via Mikrophon, denn dieses ist abhängig von der genutzten Spracherkennung. Für diese Anwendung wurde die kostenlose Spracherkennung von Google als Möglichkeit gesehen, auf einfache Art das Gesprochene in Schrift zu transkribieren. Sie erkennt Wörter der deutschen Sprache, jedoch keine einzelnen Laute oder Silben. In diesem Fall müsste auf komplexere Software zurückgegriffen werden, die Laute auf Basis ihrer Frequenzbereiche spektral analysiert und auf spezielle Wortschätze einbindet. Selbst dann könnten allerdings Störgeräusche oder Sprachfehler nicht ausgeschlossen werden. Anforderungen an ein solches System beschreiben Kröger et al. in [KBHM10], in dem die Visualisierung des Sprachsignals als Spektrum mit dem Spektrum von korrekten Artikulationen verglichen werden kann. Bei Engwall et al. [EBOK06] werden statistische Methoden genutzt. Für diese Anwendung war die automatische Spracherkennung jedoch keine zentrale Problemstellung. Eher wurde der Ansatz verfolgt, dem Anwender Laute aus einem erkannten Wort markieren und abspielen lassen zu können und ihm damit indirekte Rückmeldung zu geben. Ein anderes Feedback erhält er durch das Abspeichern seiner Spracheingabe, die jederzeit wieder angehört werden kann. So wird gleichzeitig der Fortschritt seiner Artikulation überwacht.

Einschränkungen liegen bei dieser Funktionalität auch in der Koartikulation, die als Grundlage für sponates Sprechen geübt wird. Bei der Aussprache eines Wortes wird bereits eine Eigenschaft eines Lautes, das sich im Wort befindet, am Anfang realisiert. So werden die Lippen bei Aussprache des Wortes ‚Glück‘ schon beim Anfangsbuchstaben gerundet, um den späteren Buchstaben ‚ü‘ zu bilden. Bei ‚Glanz‘ nehmen die Lippen bereits die Stellung des Vokals ‚a‘ ein. Wenn ein eingesprochenes Wort von der

Spracherkennung identifiziert wurde und alle Laute des Wortes nacheinander abgespielt werden sollen, ist die Koartikulation durch die Interpolation zwischen den Lauten begrenzt gegeben. D. h. bei jedem Buchstaben nehmen die Modelle genau die Parameter ein wie vorgeschrieben, so als würde jeder Laut einzeln ausgesprochen werden, was beim Sprechen jedoch nie geschieht. Es müsste hier untersucht werden, ob die Wahl einer anderen Interpolationsmethode die Koartikulation annähern kann, was eher ein Problemfeld der Sprachsynthese darstellt. Bei *Baldi* von Massaro [Mas06] werden für jedes Sprachsegment Gewichtsfunktionen verwendet, die den Einfluss der Zielpositionen aller beteiligten Organe bestimmen, in Relation zu den Zielpositionen seiner Nachbarsegmente. In Blender könnte es jedoch schon reichen, wenn die Bewegungen durch eine schnellere Abspielgeschwindigkeit „flüssiger“ verlaufen würden. Dann allerdings wäre auch eine Audioausgabe sinnvoll.

Wie sich aus dem Experteninterview ergab, stellte gerade die Visualisierung des Schluckvorgangs einen Vorzug der Anwendung dar, da dieser weder eingesehen werden, noch am eigenen (offenen) Mund vorgeführt werden kann. Vor allem die Zungenbewegung, die als einzige willkürlich beeinflusst werden kann, spielt dabei eine große Rolle und wurde hier anatomisch korrekt umgesetzt.

Das Mock-Up enthielt zunächst noch die Funktion, mithilfe eines Buttons die Webcam zu aktivieren um damit ein Videobild des Patienten in Form eines Spiegels zu erhalten, wurde zu anfangs zwar implementiert, jedoch wieder verworfen. Es wurde schnell offensichtlich, dass sie im Vergleich zu dem in der Therapiestunde genutzten herkömmlichen Spiegel keinen Mehrwert darstellte. Unter schlechten Lichtverhältnissen und je nach Kamera war das Videobild von deutlich schlechterer Qualität, sodass die Lippen- und Zungenstellung nicht besser einsehbar waren als bei einem Spiegelbild. Die Funktion der Webcam wurde stattdessen im zweiten Prototypen durch eine Render-Funktion ersetzt, die in Blender in die Render-Engine Cycles wechselt um das Modell qualitativ hochwertig erscheinen zu lassen. Diese Funktion wurde jedoch im Experteninterview nicht vom Logopäden beurteilt. Die Bewertung dieser Funktion hängt davon ab, wie ein oft ein Logopäde die Möglichkeit benötigen würde, bestimmte Zielpositionen der Organe auszudrucken. Für die Patienten wäre das sicherlich von Vorteil, wenn sie bestimmte Stellungen nochmals zu Hause üben könnten. Interessant wären diesbezüglich auch Videos der Animationen, die sich auf gleiche Weise über die Render-Engine generieren lassen. Fraglich bleibt jedoch, ob das in der Qualität geboten werden muss, wie sie Cycles ausgibt.

Die Umsetzung der Anwendung in Blender erweist sich als insofern nützlich, als dass sich schnell Anpassungen in der Bedienoberfläche vornehmen und sich die Visualisierung erweitern lassen. Wenn ein neuer Laut, z. B. einer anderen Sprache, hinzugefügt werden soll, wird er zunächst als neuer Schlüssel im Wörterbuch-Datentyp definiert.

Anschließend wird die Zunge über genannte Modifikatoren und Vertex-Gruppen in die Zielposition gebracht und die beteiligten Organe in die Zielstellung transformiert. Diese Parameter werden dann in Shape Keys festgehalten und für das Abspielen der Animation nur noch abgerufen.

Kapitel 8

Fazit

In der vorliegenden Arbeit wurde eine prototypische Anwendung in zwei Phasen vorgestellt, dessen interaktive 3D-Visualisierung – im Gegensatz zu patientenseitigen Spieleanwendungen oder computeranimierten Tutoren – im erster Linie dem Logopäden eine Unterstützung für die Sprechtherapie bietet. Die Umsetzung gelang mit frei verfügbarer Software und unter Einbeziehung therapeutischer Anforderungen.

Die Anwendung verwendet anatomische Modelle der oro-pharyngo-laryngealen Region, mit denen der Therapeut die Artikulation deutscher Laute in Form von Animationen in Echtzeit vorführen lassen kann. Dabei kann er auch Artikulationsorgane ausblenden lassen, um eine freie Sicht auf die Bewegungsvorgänge zu erhalten, und die Geschwindigkeit der Animation anpassen. Das Modell eines Kopfes lässt nicht nur das Formen bilabialer Laute zu, sondern auch das Vorstülpen der Lippen. Der weiche Gaumen wird bei der Lautbildung anatomisch korrekt mitbewegt. Der Phonationsstrom kann auf drei verschiedene Arten visualisiert werden und bildet die Intensität des artikulierten Lautes ab. Eine Tonausgabe auf Silbenebene dient dem Patienten als Hilfe für das Anbahnen eines Lautes, über eine Mikrofon-Anbindung kann er Worte einsprechen und abspielen lassen. Er erhält indirekt Rückmeldung über seinen Fortschritt, da er das Eingesprochene jederzeit wieder abhören kann. Es wurde gezeigt, dass aus logopädischer Betrachtung alle Funktionalitäten zusammen die Visualisierung der Zungenstellung mehr verdeutlichen als zweidimensionale Bildmaterialien oder das eigene Mundbild.

Der Therapeut kann über die Anwendung die Bewegungen eines korrekten Schluckvorgangs abspielen lassen, was dem Behandeln von pathologischen Schluckmustern und gleichzeitig als Grundlage für Übungen für die Stärkung der Zungenmuskulatur bietet.

Es wurde eine Möglichkeit vorgestellt, die Zunge über die Lautbildung hinaus in frei wählbare Positionen zu bewegen. Dafür wurde ein einfacher Kollisionstest implementiert, der die Zunge nicht mit den Zähnen des Modells zusammenstoßen lässt. Es ergab sich, dass die aktuelle Umsetzung keine komplexeren Bewegungen der Zunge für eine myofunktionelle Therapie zulässt.

Ingesamt lässt sich feststellen, dass die Anwendung durch die verschiedenen Funktionalitäten bei der Behandlung von unterschiedlichen Störungsbildern eingesetzt werden könnte.

Kapitel 9

Ausblick

Aus den Ergebnissen der vorliegenden Arbeit lassen sich Ansätze ableiten, wie die Anwendung im Hinblick auf ihre Funktionalität erweitert werden kann. Es ergibt sich, dass insbesondere bei der Bewegungssteuerung der Zunge durch den Logopäden Potential für Verbesserungen besteht. Um diese Funktionalität einsatzfähig zu machen, müsste ein erweiterter Kollisionstest greifen, der ein Zusammenstoßen zwischen Zunge und Mundraum bzw. Lippen vermeidet. Statt einer Bedienung über Schieberegler könnte dann eine 3D-Maus eingebunden werden, die die Bewegung der Zunge steuert. In Blender existieren dafür bereits Plugins. Eine solche Funktionalität könnte dann für myofunktielle Übungen eingesetzt werden, in diesem Zusammenhang müsste jedoch näher untersucht werden, ob wie hoch der tatsächliche Nutzen für die Therapie ist. In der Arbeit wurde experimentell die Funktion umgesetzt, die während der Lautbildung eine auditive Ausgabe unterstützt. Diese müsste zum einen synchron zur Animation erfolgen und zum anderen die restlichen Vokale auf Silbenebene einbinden. Ebenso könnten noch einige Verbesserungen in der Spracherkennung erzielt werden, da bisher noch keine direkte Rückmeldung über fehlgebildete Laute geboten wird. Kröger et al. [KBHM10] präsentieren diesbezüglich einen Ansatz, der das Spektrum des eingegeben Sprachsignals miteinbezieht, was sich visuell mit dem des korrekten Lauts vergleichen lässt. Quelloffene Programme wie PRAAT¹ werden bereits für Sprachanalysen in der Logopädie eingesetzt. Eine ähnliche visuelle Rückmeldung könnte sicherlich in Blender umgesetzt werden. Eine weitere, im bisherigen Prototypen ungelöste Problematik, betrifft die korrekte Koartikulation von Lauten in Sprachgebilden. Hier müsste jeder Phon in Relation zu seiner Umgebung im Sprachgefüge eine Gewichtung bekommen, die die Zielpositionen der Modelle bestimmt. Die Regeln für eine solche Funktionen müssten jedoch erst im Kontext der Sprachsynthese wissenschaftlich untersucht werden.

¹PRAAT, <http://www.fon.hum.uva.nl/praat/> (letzter Aufruf am 09.11.2014)

Andere Überlegungen, die in dieser Arbeit aufgrund der vorgegebenen Zeit in der Hintergrund gerückt sind, betreffen die Beziehung zwischen Therapeut und Patient. So könnte die Anwendung nicht nur für Logopäden, sondern auch für eine Bedienung seitens der Patienten ausgebaut werden. Speziell für Kinder könnten hierzu spielerische Konzepte mit in die Anwendung eingebunden werden. Vor allem das Kopfmodell könnte so abgeändert werden, dass das Kind sich damit identifizieren kann und zusätzlich motiviert wird. Das Modell könnte dann abhängig vom Patienten geladen werden. Um eine leichte Bedienbarkeit auch unter älteren Patienten zu erreichen, müssten bessere Interaktionswerkzeuge in Betracht gezogen werden, die über eine bisherige Maus-Bedienung hinaus gehen. Für Patienten in der neurologischen Rehabilitation könnten Eye-Tracking-Verfahren eingesetzt werden, die die Bewegungen und Lautausgabe steuern. Die Anwendung könnte ebenso für den Einsatz auf einem Tabletcomputer ausgebaut werden, sodass sie sich mobil und über einen Touchscreen bedienen lässt. Zur Messung des Therapieerfolges dieser Anwendung müsste jedoch eine vergleichende, systematische Studie erfolgen, die unterschiedliche Störungsbilder berücksichtigt. Weiterhin müssten Untersuchungen mit einbezogen werden, in denen Computertutoren den Therapeuten gegenübergestellt werden und Effektivität der sog. Visible Speech verglichen wird. Diesbezüglich gibt Massaro [Mas06] bereits einen Einblick.

Anhang A

A.1 Fragebogen zum ersten Prototypen

Fragebogen zu Prototyp I

Der Fragebogen gliedert sich in zwei Teile. Im ersten Teil soll ihre Erfahrung mit dem Computer eingeschätzt werden. Bitte füllen Sie deshalb zunächst den ersten Teil des Fragebogens aus. Die vorliegende Anwendung wird Ihnen dann kurz erläutert, bevor sie sie selbst testen können. Abschließend füllen Sie bitte den zweiten Teil des Fragebogens aus.

Die Auswertung der Untersuchung erfolgt anonym.
Personenbezogene Daten werden vertraulich behandelt.

Allgemeines

1. Einrichtung/Klinik/Praxis

.....

2. Welche Störungsbilder behandeln Sie überwiegend?

.....

Fragebogen Teil I: Umgang mit dem Computer

3. 1. Wie gut schätzen Sie Ihre Erfahrung im Umgang mit dem Computer im Allgemeinen auf einer Skala von 1 (sehr schlecht) bis 5 (sehr gut) ein?

Markieren Sie nur ein Oval.

1	2	3	4	5	
sehr schlecht	<input type="radio"/> sehr gut				

4. 2. Welche digitalen Anwendungen für die Sprachtherapie sind Ihnen bekannt?

.....

5. 3. Werden in Ihrer Einrichtung bereits digitale Anwendungen für die Sprachtherapie eingesetzt? Wenn ja, warum? Wenn nein, warum nicht?

.....

6. ----- 3.1. Falls ja, auf welcher Hardware läuft sie?

(Mehrfach-Antworten möglich)

Wählen Sie alle zutreffenden Antworten aus.

- Computer
- Smartphone
- Tablet
- Sonstiges:

7. ----- 3.2. Falls ja, welche Wahrnehmungstypen werden angesprochen?

(Mehrfach-Antworten möglich)

Wählen Sie alle zutreffenden Antworten aus.

- auditiv
- visuell
- Sonstiges:

8. ----- 3.3. Falls ja, welche Funktionen bietet die Anwendung? Name(n) der Anwendung(en)?

(Mehrfach-Antworten möglich)

Wählen Sie alle zutreffenden Antworten aus.

- Übungsmaterial in Form eines Spiels (unterhaltsames Lernen)
- Übungsmaterial in Form einer sprachtherapeutischen Geschichte (unterhaltsames Lernen)
- Übungsmaterial ohne Unterhaltung
- Zuordnung von Bildmaterial zu Buchstaben/Begriffen/Sätzen
- Zuordnung von Buchstaben/Begriffen/Sätzen zu Bildmaterial
- Lese-Sinn verstehen
- Sätze vervollständigen
- Schriftproduktion
- Videosequenzen mit auditiver Vorgabe von Lauten/Begriffen
- Abfragen des auditiven Sprachverständnisses
- Übungen zum Nachsprechen
- Aufzeichnung und Wiedergabe von Lautsprache (mit Mikrophon und Lautsprecher)
- Auditives Feedback des Nachgesprochenen
- statistische Auswertung
- Sonstiges:

9. 4. Arbeiten Sie selbst mit digitalen Anwendungen während einer Therapiesitzung?

Wählen Sie alle zutreffenden Antworten aus.

- Ja
- Nein (weiter mit Frage 5.)

10. ----- 4.1. Falls ja, mit welcher?

.....

11. ----- **4.2. Falls ja, welchen Anteil besitzt die Anwendung/Anwendungen für die Therapie? Wie viel Zeit wird darauf verwendet (z.B. in Stunden pro Monat)?**

Markieren Sie nur ein Oval.

	1	2	3	4	5	
sehr geringen Anteil	<input type="radio"/>	die Therapie besteht überwiegend daraus				

12. ----- **4.3. Falls ja, wie gut lässt sich die Anwendung Ihrer Meinung nach bedienen?**

Markieren Sie nur ein Oval.

	1	2	3	4	5	
sehr schlecht	<input type="radio"/>	sehr gut				

13. ----- **4.4. Falls ja, welche Aspekte oder welche Funktionen der Anwendung empfinden Sie als am hilfreichsten für Ihre Therapiesitzung?**

.....

.....

.....

.....

.....

14. **5. Würden Sie eine Computer-Anwendung als unterstützend für Ihre Therapie empfinden?**

Wählen Sie alle zutreffenden Antworten aus.

Ja

Nein

15. **6. Warum?**

.....

.....

.....

.....

.....

16. **7. Welche Anforderungen hätten Sie an die Anwendung?**

.....

.....

.....

.....

.....

-- Testen der Anwendung --**Fragebogen Teil II: Anwendung**

17. **1. Wenn man einem Patienten eine bestimmte Zungenstellung an sich selbst vorführen möchte, kann der Patient die Zunge nicht richtig einsehen. Wird dieses Problem bisher umgangen? Und wenn ja, wie?**

.....

.....

.....

.....

.....

18. **2. Wie häufig kommt es vor, dass Sie dem Patienten eine Zungenstellung in der Therapie-Übung verdeutlichen zu müssen?**

Markieren Sie nur ein Oval.

1	2	3	4	5	
nie	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	immer

19. **3. Welche der in der Anwendung vorgestellten Funktionen wären Ihnen für die Therapie-Sitzung praktisch am wichtigsten?**

(Einfach-Antwort)

Wählen Sie alle zutreffenden Antworten aus.

- Vorausgewählte Laute abspielen können
- Laute über das Mikrofon eingeben zu können
- Die Zunge mithilfe der Maus selbst bewegen zu können
- Die Muskeln bei einer freien Bewegung mit einbeziehen zu können
- Einzelne Körperteile ein- und ausblenden zu können
- Keine

20. **4. Auf welche der in der Anwendung vorgestellten Funktionen könnten Sie verzichten?**

(Mehrfach-Antwort möglich)

Wählen Sie alle zutreffenden Antworten aus.

- Vorausgewählte Laute abspielen können
- Laute über das Mikrofon eingeben zu können
- Die Zunge mithilfe der Maus selbst bewegen zu können
- Die Muskeln bei einer freien Bewegung mit einbeziehen zu können
- Einzelne Körperteile ein- und ausblenden zu können
- Keine

21. **5. Welchen Wert würden Sie dabei auf anatomische Korrektheit legen?**

(Mehrfach-Antworten möglich)

Wählen Sie alle zutreffenden Antworten aus.

- Einen hohen. Die Zungenstellung muss anatomisch exakt vorgeführt sein
- Einen niedrigen. Die Zungenstellung kann mit kleinen anatomischen Abweichungen vorgeführt sein, solange es realistisch aussieht

22. **6. Als wie unterstützend würden Sie diese Computer-Anwendung insgesamt für Ihre Therapie empfinden?***Markieren Sie nur ein Oval.*

1	2	3	4	5	
nicht hilfreich	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	sehr unterstützend

23. ----- **6.1. Welche Gründe sprächen Ihrer Meinung nach dafür/dagegen?**

.....

.....

.....

.....

.....

24. **7. Welche zusätzlichen Funktionen würden Sie sich für diese Anwendung wünschen? Wo liegen Verbesserungsmöglichkeiten? Haben Sie sonstige Anmerkungen?**

.....

.....

.....

.....

.....

Bereitgestellt von



A.2 Aussprache deutscher Laute

Vokal	Aussprache	Konsonant	Aussprache
[i]	Igel	[p]	Pass
[y]	hüten	[b]	Biene
[ɪ]	bitte	[t]	Tasse
[ʏ]	Hütte	[d]	Dieb
[e]	Tee	[k]	Kamel
[ø]	schön	[g]	Gast
[ɛ]	Bett	[m]	Mann
[ɛ:]	zählen	[n]	Nase
[œ]	Hölle	[ŋ]	Engel
[a]	Vase	[f]	Fell
[a:]	Fall	[v]	Wald
[ə]	Glocke	[s]	Wasser
[u]	Schule	[z]	Sonne
[ʊ]	Butter	[ʃ]	Schal
[o]	Ofen	[ç]	stechen
[ɔ]	Schloss	[j]	Ja
		[x]	suchen
		[ʁ]	Ruhe
		[l]	Los
		[r]	raus
		[ʀ]	raus

Literaturverzeichnis

- [BBPP07] BUCHAILLARD, Stéphanie ; BRIX, Muriel ; PERRIER, Pascal ; PAYAN, Yohan: Simulations of the consequences of tongue surgery on tongue mobility: Implications for speech production in post-surgery conditions. In: *The international journal of medical robotics and computer assisted surgery (MRCAS)* 3 (2007), Nr. 3, S. 252–261
- [Beł13] BĘŁTKIEWICZ, Dorota: A New Horizon in Logopaedics: Speech Therapeutic Story – Innovative Use of a Story in the Therapy of Children Speech Impediments. In: *Procedia - Social and Behavioral Sciences* 106 (2013), Nr. 0, S. 149 – 163. – ISSN 1877–0428. – 4th International Conference on New Horizons in Education
- [BSM14] BARTOLOME, G. (Hrsg.) ; SCHRÖTER-MORASCH, H. (Hrsg.): *Schluckstörungen. Diagnostik und Rehabilitation*. 5. Elsevier Verlag, 2014. – 496 S. – ISBN 978–3–437–44416–6
- [EBOK06] ENGWALL, Olov ; BÄLTER, Olle ; ÖSTER, Anne-Marie ; KJELLSTRÖM, Hedvig: Designing the user interface of the computerbased speech training system ARTUR based on early user tests. In: *Journal of Behaviour & Information Technology* 25 (2006), Nr. 4, S. 353–365
- [End99] ENDTINGER, Silke: Vorteile und Gefahren eines Computereinsatzes in der Sprachtherapie mit Kindern. In: *SAL-Bulletin. Schweizerische Arbeitsgemeinschaft für Logopädie* (1999), Nr. 94
- [Fiu10] FIUKOWSKI, Heinz: *Sprecherzieherisches Elementarbuch*. 8. Walter de Gruyter Verlag, 2010. – 507 S. – ISBN 978–3–11–023373–5
- [Gro12] GROHNFELDT, Manfred: *Grundlagen der Sprachtherapie und Logopädie*. Rheinhardt Verlag, 2012. – 214 S. – ISBN 978–3–497–02273–1
- [Gro14] GROHNFELDT, Manfred: *Grundwissen der Sprachheilpädagogik und Sprachtherapie*. Kohlhammer Verlag, 2014. – 450 S. – ISBN 978–3–17–022683–8

- [INS12] ILIE, Mihai D. ; NEGRESCU, Cristian ; STANOMIR, Dumitru: An Efficient Parametric Model for Real-Time 3D Tongue Skeletal Animation. In: *9th International Conference on Communications (COMM)* (2012), S. 129–132
- [KBHM10] KRÖGER, Bernd J. ; BIRKHOLZ, Peter ; HOFFMANN, Rüdiger ; MENG, Helen: Audiovisual Tools for Phonetic and Articulatory Visualization in Computer-aided Pronunciation Training. In: *Proceedings of the Second International Conference on Development of Multimodal Interfaces: Active Listening and Synchrony*. Berlin, Heidelberg : Springer-Verlag, 2010 (COST'09). – ISBN 3–642–12396–1, 978–3–642–12396–2, S. 337–345
- [KP01] KING, Scott A. ; PARENT, Richard E.: A Parametric Tongue Model for Animated Speech. In: *The Journal of Visualization and Computer Animation* 12 (2001), Nr. 3, S. 107–115
- [KP06] KOCSOR, András ; PACZOLAY, Dénes: Speech Technologies in a Computer-Aided Speech Therapy System. In: *Proceedings of the 10th International Conference on Computers Helping People with Special Needs*. Berlin, Heidelberg : Springer-Verlag, 2006 (ICCHP'06). – ISBN 3–540–36020–4, 978–3–540–36020–9, S. 615–622
- [mas]
- [Mas06] MASSARO, Dominic W.: The Psychology and Technology of Talking Heads: Applications in Language Learning. In: *Proceedings of International Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue System*, 2006, S. 183–214
- [PPB⁺11] PERRIER, Pascal ; PAYAN, Yohan ; BUCHAILLARD, Stéphanie ; NAZARI, Mohammad A. ; CHABANAS, Matthieu: Biomechanical models to study speech. In: *Faits de Langues* 37 (2011), S. 155–171
- [RGC98] RODRIGUES, M. ; GILLIES, D. ; CHARTERS, P.: Modelling and simulation of the tongue during laryngoscopy. In: *Computer Networks and ISDN Systems* 30 (1998), Nr. 20-21, S. 2037–2045
- [Sch10] SCHINDELMEISER, Jochen: *Anatomie und Physiologie für Sprachtherapeuten*. 2. Urban & Fischer Verlag, 2010. – 284 S. – ISBN 978–3–437–48071–3
- [SRH13] SCHARFF RETHFELDT, Wiebke ; HEINZELMANN, Bettina: Vergleich europäischer Standards und der deutschen Ausbildungssituation zur Primärqualifikation. In: *Forum Logopädie* 27 (2013), Nr. 1, S. 22–25

- [Sta13] STARKE, Anja: Einsatz von Tablets und Apps in der Sprachtherapie. In: *Logos. Fachzeitschrift für akademische Sprachtherapie und Logopädie* (2013), Nr. 21, S. 150
- [WT06] WILHELMS-TRICARICO, Reiner: *From Muscle Models to Tongue Models (And Back)*. November 2006. – Presentation at Haskins Laboratories
- [WTP97] WILHELMS-TRICARICO, Reiner ; PERKELL, Joseph: Biomechanical and Physiologically Based Speech Modeling. In: SANTEN, JanP.H. van (Hrsg.) ; OLIVE, JosephP. (Hrsg.) ; SPROAT, RichardW. (Hrsg.) ; HIRSCHBERG, Julia (Hrsg.): *Progress in Speech Synthesis*. Springer New York, 1997. – ISBN 978-1-4612-7328-8, S. 221-234
- [WZ11] WEINRICH, Martina ; ZEHNER, Heidrun: *Phonetische und phonologische Störungen bei Kindern. Aussprachetherapie in Bewegung*. 4. Springer-Verlag, 2011. – ISBN 978-3-642-17632-6