HOCHSCHULE MITTWEIDA
UNIVERSITY OF APPLIED SCIENCES

BELEGARBEIT

Studierende
Robert Nitzsche
Kathrin Schumacher
Péter Tóth

Psychoakustik von Sprache

Mittweida, 2019

Fakultät Medien

BELEGARBEIT

Psychoakustik von Sprache

Autoren:
Robert Nitzsche
Kathrin Schumacher
Péter Tóth

Studiengang:
Media and Acoustical Engineering

Seminargruppe:
MG16wC-B

5. Semester

Modul:
6479 Project-Acoustics II

Einreichung:
Mittweida, 11. März. 2019

Verteidigung/Bewertung:
Mittweida, 2019

I Inhaltsverzeichnis

II Abbildungsverzeichnis

III Tabellenverzeichnis

IV Symbolverzeichnis

f_g	Grundton in Hz
F_n	n-te Formantenfrequenz in Hz

V Abkürzungsverzeichnis

ABCs	Alphabets
best.	bestimmter
Bsp.	Beispiel(e)
bspw.	beispielsweise
bzw.	beziehungsweise
ca.	circa
cm	Zentimeter (Einheit)
d. h.	das heißt
dB	Dezibel (Einheit)
etc.	et cetera
Hz	Hertz (Einheit)
k	Kilo
m	milli, Meter (Einheit)
o. g.	oben genannten
s	Sekunde
sog.	sogenannte \-n \-s
vs.	versus
Z. B.	zum Beispiel
zw.	zwischen

1 Einleitung

Seit Jahrhunderten ist die Übermittlung von Informationen zwischen zwei Individuen notwendig. Zur Verständigung entwickelte sich die Sprache als Informationsträger. Dadurch ist die lautliche Form viel älter als die geschriebene. Mit der zunehmender Bedeutung der schriftlichen Kommunikation wurde es jedoch immer wichtiger auch die Sprache niederschreiben zu können, damit sie für jeden Menschen zugänglich sind. Aufgrund von technischen Entwicklungen ergaben sich neue Möglichkeiten Sprache darzustellen. So ist bspw. die „sichtbare Sprache“ entstanden. Sie wird heutzutage genutzt um Sprachmerkmale zu erkennen bzw. zu verstehen. Unteranderem keimte auch die Sprachverarbeitung auf, die z. B. zur Mensch-Maschine-Kommunikation gebraucht wird und heute noch immer nicht ausgereift ist. Aufgrund dieser Fortschritte bekam die Analyse von Sprache immer mehr an Bedeutung.
Durch die sehr große Variationsbreite und dem umfangreichen Wissensbestand ist es jedoch schwer den Überblick dieses Themas zu behalten.

Das Ziel dieser Arbeit ist eine verständliche Basis über die Merkmale und Eigenschaften von Sprache zu schaffen. Zudem soll eine Übersicht der verschiedenen Ansatzpunkte der Analyse (psychoakustisch und Phonem bedingt) dargelegt und mit Hilfe eigener Aufnahmen untersucht werden, sodass ein Gespür für dieses wichtige Kommunikationsmittel entsteht.
Die hier behandelten Themen umfassen nur einen geringen Teil und legen ausschließlich die grundlegenden Bausteine der Sprache dar. Dazu werden vorerst das Erlernen sowie die Entstehung von Sprache erläutert. Anschließend folgt eine kurze Einführung der Sprachlaute, die in einem späteren Kapitel aufgegriffen und explizit mit eigenen Beispielen erklärt werden. Zudem enthält das zweite Kapitel eine Übersicht der Realisierung von Sprachmodellen sowie eine Zusammenfassung der wesentlichen Merkmale von fließender Sprache. Zum Schluss wird noch auf einige psychoakustische Parameter eingegangen, die anhand von aufgenommenen Beispielen verdeutlicht werden.
Die Aufnahmen fanden Hochschulintern statt und wurden Mittels des Programms ArtemiS und einem Kunstkopf aufgenommen und ausgewertet.

Die entsprechenden Abschnitte wurden verfasst von:
Robert Nitzsche: Kap. 3, 5
Kathrin Schumacher: Kap. 1, 4, 6,
Péter Tóth: Kap. 2

2 Allgemeines zur menschlichen Sprache

Die Menschheit, wie jedes Lebewesen, hat ihre eigenen Mittel zur Mitteilung bestimmter lebenswichtiger Informationen, und je mehr weiterentwickelt und intelligenter das Wesen ist, umso mehr Informationen kann es mit den anderen auf unterschiedlicher Weise kommunizieren. Bei dem Mensch gibt es mehrere Möglichkeiten um die Information zu kodieren, sog. Kommunikationsmittel, wie z. B. Gestik, Mimik, Schreiben, Zeichnen oder die Sprache. Im Folgenden werden wir uns mit der Sprache näher auseinandersetzen, und zwar mit der Entstehung, dem Erlernen, der Modellvorstellung der Sprache und was bei der fließenden Sprache beachten werden muss. [Eppinger & Herter 1993, S. 3]

2.1 Entstehung der Sprache

Bei der Sprache dient die menschliche Stimme als Informationsträger. Allerdings ist das Erlernen dieser Art der Informationskodierung und -übertragung ein langer, komplizierter Vorgang. Der Mensch braucht in der Regel mehrere Jahren zum Erlernen und Aufbessern der Sprache. Dies ist ein ähnlicher Vorgang, wie das Erlernen eines Instruments, z. B. eines Blasinstruments. Erstens muss man in der Lage sein, einen hinreichend starken Luftstrom erzeugen zu können, sodass die Resonatoren des Instruments überhaupt angeregt werden und die gewünschten Frequenzen dauerhaft gehalten werden. Anschließend muss erlernen werden, wie die einzelnen Töne der gesamten Tonskala des Instruments zu erzeugen sind. Am Ende müssen die einzelnen Feinheiten, Melodien, Lautstärkeänderungen, etc. geübt und aufgebessert werden, bis das Musikinstrument gut beherrscht wird. Das Erlernen der Sprache fängt bei dem Mensch schon ab dem Alter von ca. zwei Monaten an und dauert bis zu dem siebten oder achten Lebensjahr. Wenn angenommen wird, dass das Kind gesund ist und von den Eltern richtig gelehrt bzw. motiviert wurde, sieht der Vorgang wie folgt aus:

- ca. ab dem zweiten Monat fängt das Kind an, unterschiedliche Töne auszugeben und zu „blabbern“.
- Ab dem Alter von fünf Monaten „blabbert“ das Kind einzelne Silben.
- Im ersten Lebensjahr werden vom Kind auch sog. Einwortsätze verwendet, die schon ihre eigene Bedeutung tragen. Das schnellste Entwicklungsstadium des Lernvorgangs findet etwa zwischen dem 12. und 13. Monat statt.
- Am Ende des zweiten Lebensjahrs kann das Kind einfache Sätze bauen und hat einen Wortschatz von ca. 200 Wörtern, der überwiegend aus Substantiven und Verben besteht.
- Bis zum vierten Lebensjahr kann das Kind sich verständlich und deutlich ausdrücken, komplexere Sätze bauen und Wörter vielfältiger Wortarten

anwenden. Die meisten Laute werden beim spontanen Reden fehlerfrei ausgesprochen (ausgenommen „sch“ und „r“).

- Bis zum Beginn der Grundschule kann das Kind, das reif genug ist und dessen Entwicklung in Ordnung war, sich fließend (ohne Pausen oder Wiederholungen) mit fehlerfreier Aussprache ausdrücken. Sein Wortschatz ermöglicht es ihm, anhand eines Bildes eine Geschichte zu erzählen. [Fodor o. D.]

Für die Entstehung der menschlichen Sprache ist kein einzelnes Organ zuständig, sondern mehrere Organe und deren fein abgestimmtes Zusammenspiel. Somit wird nicht von einem „Stimmorgan“ sondern einem „Stimmapparat“ gesprochen. Erstens wird ein Luftstrom von der Lunge geliefert, der durch den Kehlkopf und die Stimmritze strömt bzw. gezwungen wird. Die Stimmritze wird zu Schwingungen angeregt, die den Luftstrom in ein regelmäßiges Anregungssignal verwandelt (**Abb. 1**).

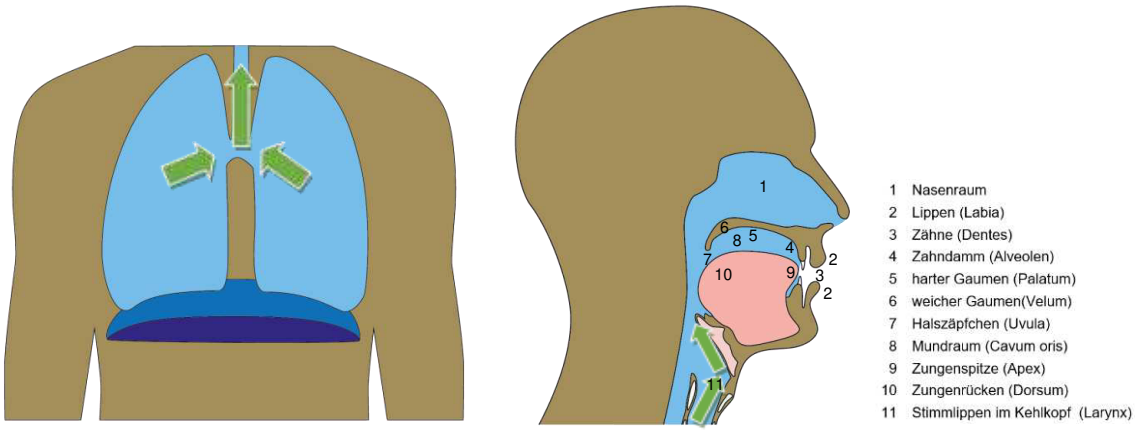


Abb. 1: Der Weg des Luftstroms von der Lunge bis zum Kehlkopf, eigene Darstellung (R. Nietzsche) nach [Pfister & Kaufmann 2017, S. 13] entworfen

Der Luftstrom kann allerdings unterschiedlich angeregt werden abhängig davon, ob die Stimmritze geöffnet oder geschlossen ist. Je nach Art der Anregung können die entstehende Laute in zwei Kategorien eingeteilt werden: *stimmhafte* und *stimmlose* Laute. [Eppinger & Herter 1993, S. 3 f.]

2.1.1 Stimmhafte Laute

Wenn die Stimmritze geschlossen ist, d. h. die Stimmlippen sind ganz eng beieinander (**Abb. 2**), kann die Luft nicht einfach passieren, sondern es baut sich Überdruck auf. Wird dieser Druck zu groß, geben die Stimmbandmuskeln nach und die aufgestaute Luft kann entweichen. Die Stimmbänder schließen sich nun wieder und der Vorgang wiederholt sich. Somit entsteht ein periodisches, sägezahnähnliches Signal, dass das Anregungssignal für stimmhafte Laute ist.
Dieses Anregungssignal liefert auch die Grundfrequenz f_g der Sprache: je schneller bzw. öfter sich die Stimmbänder schließen und wieder öffnen, desto höher ist diese Frequenz. Die Grundfrequenz steht außerdem mit der Höhe der menschlichen Stimme und der Anspannung der Stimmbänder im direkten Zusammenhang. Sie beträgt bei Kindern bis zu $330\ \mathrm{Hz}$ und kann bei Männern eine Tiefe von ca. $80\ \mathrm{Hz}$

erreichen. Lautstärke und Tonhöhe sind allerdings nicht ganz unabhängig voneinander zu beeinflussen: Lautstärke steht mit der Amplitude der Stimmbandschwingung im direkten Zusammenhang. Zu einer größeren Amplitude ist ein höherer Luftdruck nötig, wodurch die Stimmbandmuskeln mehr angespannt werden müssen. Eine größere Anspannung bedeutet aber auch eine größere Tonhöhe. Von den Sprachlauten sind alle Vokale stimmhaft.

Abb. 2: Stimmritze geschlossen

2.1.2 Stimmlose Laute

Bei der Entstehung von stimmlosen Lauten ist die Stimmritze – im Gegensatz zu dem Fall der stimmhaften Lauten – weit geöffnet, sodass der Luftstrom ohne großes Hindernis kontinuierlich durch den Kehlkopf passieren kann. Der Luftstrom wird zu Schwingungen bzw. Wirbelungen angeregt, somit entsteht ein rauschartiges, unregelmäßiges Signal. Dies dient als Anregungsfunktion für stimmlose Laute. [Eppinger & Herter 1993, S. 4 ff.]

2.1.3 Die Bildung von Sprachlauten

Nachdem der Luftstrom im Kehlkopf zu einem bestimmten Anregungssignal umgewandelt wurde, gelangt er in den sog. Vokaltrakt. Unter Vokaltrakt ist das Zusammenwirken mehrerer Organe zu verstehen, die als Hohlraumresonatoren funktionieren und die Bildung von den unterschiedlichsten Sprachlauten ermöglichen. Zu den Hauptresonatoren gehören Rachenraum, Nasen- und Mundhöhle, die in ihrer Form und Größe willentlich veränderbar sind. Einen weiteren Beitrag bei der Stimmbildung leistet der Atemtrakt. Dazu zählen die Luftröhre, die Bronchien, der Kehlkopf und die Nasennebenhöhlen. Im Gegensatz zu den Hauptresonatoren sind diese in ihrer Form bzw. Größe nicht willentlich veränderbar und somit für die sprechertypischen Merkmale der Sprache zuständig. Atemwegserkrankungen oder bronchiale Infekte haben allerdings auch einen Einfluss darauf. Bei bspw. einer Erkältung wird die Stimme geändert. Das Frequenzspektrum des Anregungssignals wird vom Atemtrakt beeinflusst. Es werden bestimmte Frequenzbereiche dabei durchgelassen und andere unterdrückt. Die durchgelassenen Frequenzbereiche werden auch *Formanten* genannt, die bei der Bildung von Vokalen besonders deutlich sind.

Durch die Beweglichkeit und Vielzahl der relevanten Organe ergeben sich unendlich viele erzeugbare Laute, wovon nur eine kleine Untermenge für die Verständigung angewendet wird, die sog. *Phoneme* oder *Sprachlaute*. Phoneme haben keine eigene Bedeutung und werden in Lautschrift dargestellt. Die meisten Sprachen bestehen aus $30 \ldots 50$ Phonemen. Sprachen werden allerdings schwer durch Phoneme beschrieben, denn ein Phonem kann abhängig von Sprecher, Betonung,

umgebenden Sprachlauten, etc. unterschiedlich ausgesprochen werden. Sprachlauten können in zwei Gruppen eingeteilt werden: Vokale und Konsonanten. [Eppinger & Herter 1993, S. 6 f.]

2.2 Modellvorstellung der Spracherzeugung

Mit Kenntnissen über die menschliche Spracherzeugung bzw. über Signalgeneratoren, Filter und Verstärker kann man sich überlegen, wie ein künstliches, elektrisches Sprachmodell erstellt werden kann.
Im Folgenden werden die Umsetzung und der grobe Aufbau solcher Systeme betrachtet.

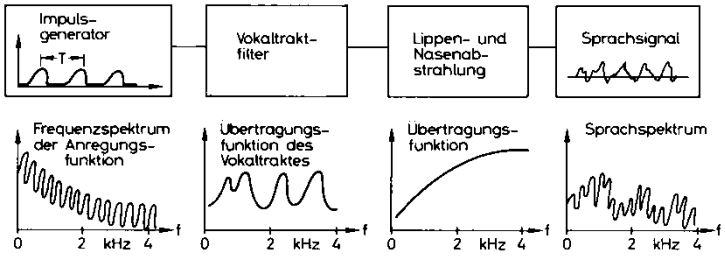


Abb. 3: Grober Aufbau eines Sprachmodells, [Eppinger & Herter 1993, S. 12]

Um das Modell aufbauen zu können, müssen als erstes die relevanten Organe, je nach ihrer Funktion bei der Spracherzeugung, mit jeweils einem passenden Bauteil ersetzen werden (**Abb. 3**). Der Kehlkopf bzw. die Stimmbänder dienen dabei als Signalquelle, der gesamte Vokaltrakt als unterschiedliche Filter bzw. Übertragung, und die Abstrahlung von Mund- und Nasenhöhlen als Verstärker. Um das Modell einfacher zu gestalten werden Rückkopplungen zum Gehirn bzw. Gehör, sowie Verbindungen zwischen Lautanregung und Resonanzräume ober- und unterhalb der Stimmbändern nicht berücksichtigt. [Eppinger & Herter 1993, S. 11 f.]

2.2.1 Theorie des Linearen Filters

Das lineare Filter besteht im Wesentlichen aus den oben genannten Bauelementen. Eine Anregungsquelle erfüllt die Funktion des Kehlkopfs bzw. der Stimmbänder sowie unterschiedliche Filter für den Vokaltrakt und ein Verstärker für die Abstrahlung. Die Bauteile sind unabhängig, d. h. ohne Kopplung.

Als Quelle soll ein Impuls- und ein Rauschgenerator eingesetzt werden. Ein Impulsgenerator mit einem ungefähr sägezahnförmigen Signal für die stimmhafte und ein Rauschgenerator für die stimmlose Anregung. Bei der stimmhaften Anregung beträgt der Abstand der Spektrallinien im Frequenzbereich ca. die der Grundfrequenz. Je tiefer die Grundfrequenz ist, umso dichter wird das Spektrum. Die Einhüllende sollte dabei eine kontinuierliche Abnahme mit zunehmender Frequenz haben: $12\ \mathrm{dB}$ pro Oktave. Da einige Laute gleichzeitig eine stimmlose und eine stimmhafte Anregung haben, ist es nützlicher, zwischen den Quellen einen Mischer statt eines Umschalters einzusetzen.

Als Filter werden zur Vereinfachung lineare Filter eingesetzt, obwohl der Artikulationstrakt ein nichtlineares Verhalten aufweist. Z. B. sind für Vokale drei bis fünf Formanten nötig, um einzelne Vokale voneinander unterscheiden zu können bzw. um ein Sprachsignal guter Qualität zu erhalten. Das heißt, hier müssten mehrere Bandpassfilter zum Einsatz kommen. Die einzelnen Filter können entweder Parallel oder in Reihe angeordnet werden, diese Anordnungen haben allerdings Vor- und Nachteile. Bei der Parallelanordnung (**Abb. 5**) arbeiten die Filter voneinander unabhängig, die Amplituden der Filterdurchlasskurven müssen aber getrennt eingestellt werden. Diese Amplitude ist bei der Reihenanordnung (**Abb. 4**) nur einmal einzustellen. Die Filter können allerdings nicht mehr voneinander unabhängig arbeiten.
Ein weiteres Problem besteht darin, dass die gesamten Filter vor der Erstellung jedes Phonems neu eingestellt werden müssen, d. h. „fließende Sprache“ zu erstellen wäre sehr aufwändig und eventuell auch nicht „fließend“ realisierbar. In der konventionellen Technik werden dafür sog. Switched-Capacitor-Filter oder SC-Filter eingesetzt, die eine variabel einstellbare Frequenz und Bandbreite besitzen. [Eppinger & Herter 1993, S. 13 ff.]

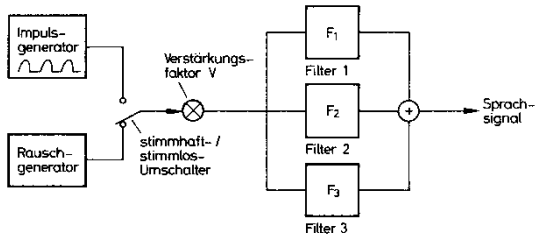


Abb. 5: Parallelanordnung von Filtern, [Eppinger & Herter 1993, S. 14]

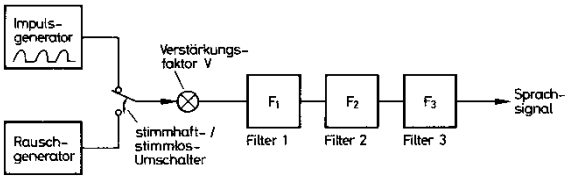


Abb. 4: Reihenanordnung von Filtern, [Eppinger & Herter 1993, S. 14]

2.3 Fließende Sprache

2.3.1 Koartikulation und Prosodie

Vorhin haben wir uns mit der Bildung von Sprachlauten beschäftigt. Bei der fließenden Sprache reicht es allerdings nicht aus, nur über die Bildung einzelner Phoneme Bescheid zu wissen: Wörter sind nicht nur Aneinanderreihungen von Phonemen, sowie Sätze auch nicht Aneinanderreihung von Wörtern sind. Eine große Rolle spielt die sog. *Koartikulation*. Koartikulation ist der Einfluss von benachbarten Phonemen bzw. Wörtern aufeinander. Ein Beispiel dafür ist die Änderung der Stimmhaft- bzw. Stimmlosigkeit bei Konsonanten: im Wort „abstrakt“ stehen das stimmhafte „b“ und das stimmlose „s“ nebeneinander, aber bei der Aussprache ist der zweite Konsonant „dominant“ und beeinflusst damit die Bildung des Ersten. In unserem Wort wird das „b“ zu einem „p“, d. h. statt „abstrakt“ sprechen wir „apstrakt“ aus.

Die anderen, lautlichen Eigenschaften der fließenden Sprache bezüglich Betonung, Grundfrequenz, Tempo, Rhythmus, etc. umfasst der Begriff „Prosodie“. Diese Eigenschaften beeinflussen in erster Linie die physikalischen Eigenschaften des Signals. Die Grundfrequenz ist dabei allerdings nicht nur von der Stimme des

Sprechers abhängig: durch die Veränderung der Grundfrequenz bzw. deren Verlaufs kann ausgedrückt werden, ob es sich bei einem gesprochenen Satz um eine Frage, eine Aussage oder einen Befehl handelt. Bei den meisten Fragen und Aussagen sind die folgenden einfachen, typischen Muster zu erkennen: Bei Fragen wird die Grundfrequenz am Ende des Satzes oder sogar im ganzen Satz angehoben; bei Aussagen sinkt diese. Bei Entscheidungsfragen im Ungarischen übernehmen diese Frequenzverläufe sogar die Hauptrolle. Da die Struktur des Satzes erhalten bleiben soll, kann der Zuhörer nur aus diesem Verlauf erraten, ob es sich um eine Aussage oder eine Frage handelt. Schriftlich wird dieser Unterschied alleine durch das Fragezeichen ausgedrückt. Z. B.: Die Aussage „Du kommst mit mir ins Kino." lautet auf Ungarisch „Eljössz velem a moziba.", währenddessen die dazugehörige Frage „Kommst du mit mir ins Kino?" läutet „Eljössz velem a moziba?". Man muss allerdings dabei anmerken, dass bei den Grundfrequenzverläufen nicht der absolute Wert der Frequenz wichtig ist, sondern die Änderung davon. So können die bestimmten Muster der Verläufe genauso bei hohen Kinderstimmen als auch bei Erwachsenen Frauen und Männern, trotz sehr unterschiedlicher absoluter Grundfrequenzen, erkannt werden. Außerdem gibt es für mehrere unterschiedliche Arten von Sätzen ein allgemeines Muster. Diese Muster sind im Tonsystem von Halliday näher erläutert (**Abb. 6**). Allerdings sind die o. g. Verläufe nicht ganz linear: an allen

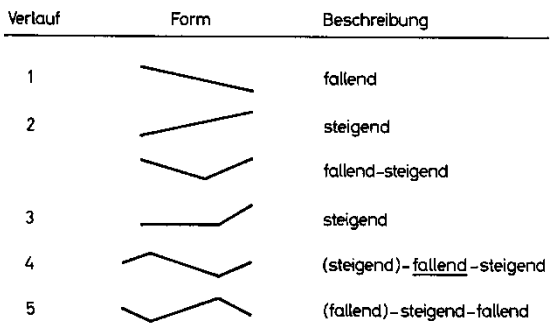


Abb. 6: Das Tonsystem von Halliday, [Eppinger & Herter 1993, S. 47]

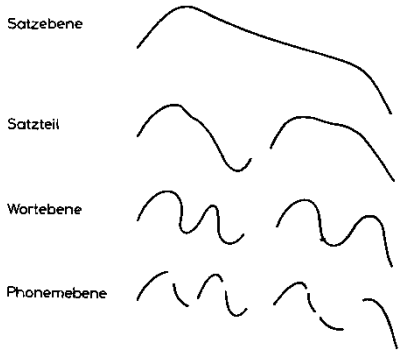


Abb. 7: Darstellung der Grundfrequenzverläufe, [Eppinger & Herter 1993, S. 36]

Sprachebenen (Satz-, Wort-, Phonemebene) sind Schwankungen zu erkennen (**Abb. 7**).

Die Betonung von Wörtern ist ein weiteres Ausdrucksmittel, üblich werden Lautstärke, Silbendauer (die betonten Teile der Sprache werden länger ausgesprochen) bzw. seltener auch Tonhöhenänderungen angewendet. Es gibt aber keine einheitliche Methode, es kann je nach Sprecher und Situation unterschiedlich sein. D. h. die Erkennung von Betonungen ist mit Sprachverarbeitungssystemen schwierig zu realisieren. Aber in der technischen Sprachverarbeitung, wo die Bedeutungsunterschiede wichtig sind (z. B. Siri), müssen Systeme in der Lage sein, solche Feinheiten der Sprache zu erkennen. Bei Systemen der Sprache-Text-Umsetzung ist dies nicht mehr nötig.

Die Sprechgeschwindigkeit ist oft an die äußeren Umständen angepasst bzw. hängt vom Sprecher ab (z. B. angespannt oder müde). Aber nicht alle Laute lassen sich

vom Tempo beeinflussen, Explosivlaute (wie p, t, g, etc.) werden bspw. davon unabhängig ausgesprochen. Das muss bei der Tempoänderung von künstlicher Sprache beachtet werden. Ein anderes Problem betrifft die Vokale: zu schnelle Aussprache kann bei Vokalen dazu führen, dass die Formantfrequenzen nicht erreicht werden, d. h. der Vokal ist nicht mehr deutlich zu identifizieren. In diesem Fall wird von *Vokalreduktion* gesprochen (**Abb. 8**).

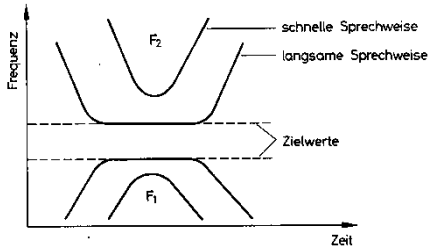


Abb. 8: Vokalreduktion bei der zu schnellen Aussprache, [Eppinger & Herter 1993, S. 37]

Relevante Größen zur Beschreibung der Sprechgeschwindigkeit sind die mittlere Sprechgeschwindigkeit und der Satzrhythmus. Die mittlere Sprecheschwindigkeit kann durch die Anzahl gesprochener Wörter über eine längere Zeit (z. B. eine Minute) beschrieben werden. Der Satzrhythmus beschreibt die Änderung der Sprechgeschwindigkeit innerhalb eines Satzes. Dies kann durch eine unterschiedliche Silbendauer zum Ausdruck kommen. Das Sprechtempo wird abhängig von der Umgebung und von der Art der Aussage ebenfalls angepasst, z. B. bei einer lyrischen Lesung oder in der Kirche ist es langsamer und in der Umgangssprache eher schneller. Dies ändert nichts an der Bedeutung, kann aber bestimmte Teile der Aussage stärken bzw. unterstützen. Somit wird das Tempo auch als rhetorisches Hilfsmittel angewendet. Es gibt allerdings Wörter, bei denen ein zu schnell ausgesprochener Vokal oder Phonem tatsächlich die Bedeutung ändert, z. B. die Wörter „Lamm“ und „lahm“ oder „Masse“ und „Maße“. Um solche Unterschiede auch für Sprachsysteme erkennbar zu machen, sollten Untersuchungen durchgeführt werden, die bestimmen ab welcher Länge des Phonems das eine oder das andere Wort wahrgenommen wird. [Eppinger & Herter 1993, S. 35 ff.]

2.3.2 Sprachliche Zusammenhänge

Auf die Verständlichkeit eines Satzes haben Wort- und Sinnzusammenhänge auch Einfluss, z. B. im Satz „Auf der Straße fährt ein Auto.“ ist das Wort „Auto“ besser zu verstehen als in „Ich sehe ein Auto.“. Der Grund dafür ist, dass zum Verb „fahren“ inhaltlich weniger Wörter passen, als zu „sehen“. Wird von grammatikalischer Korrektheit ausgegangen, können nach dem Hören eines Artikels auch einige Wörter ausgeschlossen werden, denn z. B. nach dem Artikel „das“ kommt nicht das Wort „Musik“. Diese Zusammenhänge werden in mit Wahrscheinlichkeiten arbeitenden Modellen, wie z. B. die Markov-Modelle, berücksichtigt. In der Regel werden dabei Wahrscheinlichkeiten von Wörtern, Phonemen oder Wortkombinationen aus zwei oder mehreren Wörtern untersucht. [Eppinger & Herter 1993, S.37 f.]

3 Messaufbau und -ablauf

Die Sprachaufnahmen sind unter Verwendung des Kunstkopfes Typ: HSM II.4 (Geräte Nr.: 12406009) entstanden. Die erzeugten Signale wurden mit Hilfe der "ArthemiS Suite 9.3" aufgenommen und abgespeichert sowie durch weitere Analysen ergänzt.

Da für die Untersuchung ein Monosignal vollkommen ausreichend ist, wurde entschieden, lediglich den rechten Stereo-Kanal zur Auswertung zu verwenden. Um dort ein möglichst sauberes Signal zu erhalten, wurden Bestimmungen festgelegt: Der Sprecherabstand betrug bei jeder Messung $50\ \mathrm{cm}$ zum rechten Ohr des Kunstkopfes und es wurde unter einem Winkel von 45 Grad (Horizontalebene) und in einer Höhe von $1{,}2\ \mathrm{m}$ eingesprochen. Diese Festlegungen haben sich als guter Kompromiss erwiesen.
Der Abstand, sowie Winkel garantieren einen akzeptablen Pegel, der Direktschall ist dominierend und die Membran des Mikrofons wird nicht durch den Wind der stark ausströmenden Luft zu Schwingungen angeregt.

Abb. 10: Messaufbau mit Kunstkopf, eigene Aufnahme

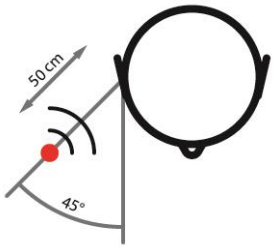


Abb. 9: Einsprechschema, eigene Darstellung (R. Nietzsche)

Insgesamt haben 4 Probanden (2 weibliche, 2 männliche) dieselben Inhalte eingesprochen. Viele Sprechertexte wurden nacheinander stimmhaft (normal gesprochen), anschließend stimmlos (geflüstert) und jeweils mit normalen, sowie langsamen Sprechtempo aufgezeichnet. Dies hat den Sinn, eine bessere Vergleichbarkeit unter den Aufzeichnungen zu garantieren und verschiedenen Forschungsfragen nachgehen zu können. Als Sprechertexte wurden die drei Kategorien: einzelne Laute und Worte, als auch ganze Sätze verwendet. Je nach Untersuchungskriterium konnten diese anschließend genauer betrachtet werden.
Die Aufnahmen wurden im Haus 39 (Raum 137) getätigt. Die Lokalität ist durch eine Vielzahl von Störfaktoren beeinflusst. Das Zentrum für Medien und Soziale Arbeit ist ein sehr beschäftigter Ort. Oft dringen Umgebungsgeräusche vom Flur und den angrenzenden Räumen in das „Aufnahmestudio“ ein. Auch Verkehrslärm ist dort eine Einflussgröße. Die Fensterfront dieses „Studios“ ist einer Hauptstraße zugewandt. Lastkraftwagen erzeugen ein spürbares Dröhnen im Raum. Diese Einflüsse können sich den Aufnahmen beimischen und eine Auswertung erschweren.

4 Sprachlaute

Sprachlaute können in zwei Lautklassen unterteilt werden: Vokale und Konsonanten. Im folgenden Kapitel werden die linguistischen sowie physikalischen Eigenschaften dieser beiden Gruppen aufgezeigt. Zudem werden grundlegende Begriffe, die in diesem Beleg Verwendung finden, vorgestellt und erläutert. In den späteren Subkapiteln befinden sich außerdem Beschreibungen zu Erkennungsmerkmalen im Spektrum und eine knappe Erklärung des Spektrogramms. Als letztes werden Analysen von einzelnen Lauten sowie eines Wortes und einer wortkette dargelegt. [DUDEN Bd. 6 2015, S. 23]

4.1 Grundsätzliches

Phon vs. Phonem

Als „*Laut*“ (*Phon*) wird jede einzelne laute Äußerung eines Lebewesens bezeichnet. Durch stetig wiederkehrendes Aufkommen der Phone in ähnlicher Form können diese in verschiedene Gruppen aufgeteilt werden. Sie sind durch die *International Phonetic Association* (*IPA*) in ein allgemeingültiges Zeichensystem eingebettet, dass übergeordnet als *Lautschrift* bezeichnet wird (vgl. Kap. 2.1.3). Zur besseren Verständlichkeit bzw. Identifikation dieser Zeichen befindet sich im Anhang eine ausführliche Tabelle der im Deutschen vorkommenden Kardinallaute.

Für die Verständigung der Menschen untereinander wird jedoch nur eine geringe Anzahl von unendlich vielen Realisationsmöglichkeiten verwendet. Diese werden dagegen als *Sprachlaute* bzw. *Phoneme* klassifiziert (vgl. 2.1.3). Sie sind die „kleinsten bedeutungsunterscheidenden Einheiten einer Sprache“ [DUDEN Bd. 6 2015, S. 20]. D. h. das zwei Phoneme in der selben sprachlichen Umgebung existieren können aber einen Kontrast zueinander bilden müssen, sodass der Austausch des einen Sprachlauts durch den Anderen zwei unterschiedlich bedeutende Wörter zur Folge hat. Um diese zwei Arten differenzieren zu können werden Phone in eckige Klammern ([…]) und Phoneme in Schrägstriche (/…/) gesetzt. Eines der Hauptprobleme bei der Darstellung durch Phoneme ist, dass Sprachlaute verschieden artikuliert werden können. Diese unterschiedlichen Aussprachevariationen ein und desselben Phonems werden als *Allophone* bezeichnet. Die größten Faktoren der Beeinflussung sind die Person und deren Verfassung (z.B. Gesundheit, Müdigkeit), die Art und Betonung des Satzes sowie die umgebenden Phoneme (vgl. Kap. 2.1.3). [Eppinger & Herter 1993, S. 6 f. & S. 30 ff.] [DUDEN Bd. 6 2015, S. 20]

Phonologie vs. Phonetik

Die linguistische Handhabung der Sprache (*Phonologie*) behandelt hauptsächlich die Bedeutung der Sprachelemente sowie deren Anordnung und Beeinflussung.

Erst durch einen sprachlichen Zusammenhang (*Kontext*) kann die Bedeutung einer Lautfolge geklärt werden. Für eine phonetische Untersuchung ist ein Sprachsystem zwingend erforderlich, wobei Umstände wie etwa das Geschlecht sowie der emotionale oder gesundheitliche Zustand des Sprechers völlig belanglos sind.
Demgegenüber befasst sich die *Phonetik* mit der Bildung und Artikulation des Lautes, also der Entstehung des Schallereignisses. Hier ist es wichtig, ob die Laute wütend oder eher ängstlich hervorgebracht werden oder der Sprecher bspw. eine Erkältung austrägt. [Eppinger & Herter 1993, S. 28 ff.]

Formanten
Die Bereiche mit stark erhöhter Spektralenergie sind die lokalen Maxima des Spektrums. Sie werden als Formantenbereiche (kurz *Formant, vgl. Kap. 2.1.3*) bezeichnet und besitzen eine ihnen zugehörige *Formantenfrequenz* F_n. Sie sind die willentlich erzeugten Resonanzfrequenzen des im Vokaltrakt geformten Signals. Zur Charakterisierung bzw. Erkennung eines Phonems können ein und/oder mehrere dieser ausgeprägten Gebiete genutzt werden. Sie besitzen keine konforme Lage im Frequenzbereich und unterscheiden sind sogar beim selben Sprecher. Die Formantenlage ist vom Geschlecht des Sprechers unabhängig.
Mit steigender Frequenz nimmt das „Informationsgewicht“ ab. Der Grund hierfür ist die geringer werdende Artikulationsabhängigkeit. Z. B. beträgt die Variationsrate der ersten Formantenfrequenz fast ±100 %, wohingegen die des vierten Formanten beinahe konstant bleibt. [Kuttruff 2004, S. 222] [Meyer & Neumann 1967, S. 228] [Pfister & Kaufmann 2017, S. 19] [Terhardt 1998, S. 195]

4.2 Eigenschaften

Die Unterteilung der Sprachlaute erfolgt in zwei Klassen: Vokale und Konsonanten (vgl. Kap. 2.1.3). Sie werden auf Grund zusätzlicher Merkmale wie bspw. stimmhafter und stimmloser Lauterzeugung, dem Artikulationsort oder der Artikulationsart weiter in Gruppen unterteilt. Dabei bedeutet stimmhaft ein Mitschwingen der Stimmlippen (periodische Grundfrequenz) und stimmlos ein Mangel dessen (vgl. Kap. 2.1.1 und Kap. 2.1.2). Der Artikulationsort ist der Bereich im Mundraum, indem der entsprechende Laut gebildet wird. Er ist abhängig von der Position (z. B. in Mundhöhle) und den beteiligten Organen (z. B. Engpass zw. Zunge und weichem Gaumen). Die Artikulationsart beschreibt die Luftführung oder -beeinflussung durch Mund- und Nasenraum.
Zu einer Phonemerkennung ist die Zerlegung eines Sprachsignals sinnvoll, weil die physikalischen Eigenschaften stark voneinander abweichen. Ein Phonem kann dank der Formantenfrequenzen als best. Vokal identifiziert werden. Die Konsonanten werden im nächsten Schritt weiter untergliedert (z. B. Nasallaute, Explosivlaute, stimmhafter oder stimmloser Frikativlaut). [DUDEN Bd. 6 2015, S. 20] [Eppinger & Herter 1993, S. 6 f., S. 30 ff.]

4.2.1 Vokale

Vokale sind Öffnungslaute. D.h. die Atemluft kann ungehindert aufgrund der freien Mittelpassage des Sprachtrakts durch Mund und/oder Nase ausströmen. Alle Vokale sind stimmhaft, sie besitzen eine periodische Grundfrequenz (vgl. Kap.2.1.1). Der Vokaltrakt ist die schallmodifizierende Komponente und darf niemals zu beträchtlichen Engstellen, z. B. durch den Zungenrücken, geformt werden. [Eppinger & Herter 1993, S. 7] [Terhardt 1998, S. 188] [Pfister & Kaufmann 2017, S. 14]

Aufgrund der Resonanzeigenschaften des Artikulationstraktes entwickeln sich Fomantenfrequenzen heraus, die für jeden Vokal spezifisch sind. Zur besseren Verständlichkeit sind in **Abb. 12** verschiedene Beispiele für die Zungenposition von Vokalen (erste Reihe) sowie eine graphische Darstellung der ins Verhältnis gesetzten artikulatorischen Parameter (**Abb. 14**) gegeben. Stellt man ein /a/ einem /i/ gegenüber, so ist erkennbar, dass sich der Zungenrücken bei /a/ hinten und weit unten befindet und der Mund eine offene Stellung aufweist. Die Position der Zunge beim /i/ hingegen ist weit vorne und der Mund nahezu geschlossen. Bei der eigenen Artikulation beider Laute hintereinander lässt sich erschließen, dass das /a/ verhältnismäßig tiefer klingt. Die Rundung der Lippen ist bei bewusster Kontrolle ebenfalls spürbar. Beide Phone sind ungerundet. Wird dagegen ein /u/ ausgesprochen sind die Lippen *labial* (gerundet).

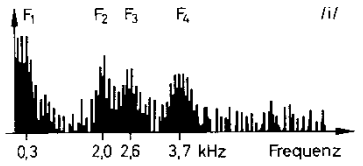


Abb. 13: Formanten der Vokale /i/ und /a/, [Eppinger & Herter 1993, S. 8]

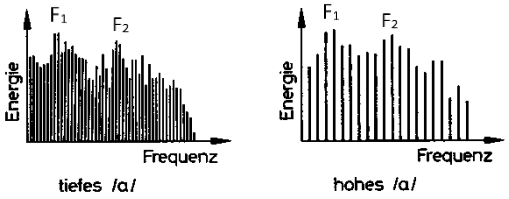


Abb. 11: Aussprache eines Vokals mit unterschiedlicher Stimmhöhe, [Eppinger & Herter 1993, S. 9]

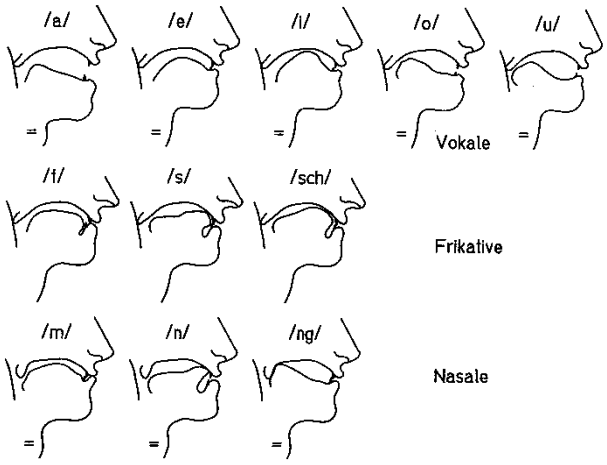


Abb. 12: Zungenstellung bei verschiedenen Lauten [Terhardt 1998, S. 184]

Abb. 13 zeigt die Bereiche der Formantenfrequenzen für den Vokal /i/. Es zeigt anschaulich, dass die Frequenzen vorerst abklingen bis sie sich in den Formantenbereichen wieder stark erhöhen. [Eppinger & Herter 1993, S. 8 f.]

In **Tab. 1** befinden sich zusätzlich die in Zahlwerten formulierten Richtwerte der Formanten der Kardinalvokale. Es lässt sich eine gewisse Korrelation zw. Mundöffnung bzw. Zungen- und Formantenlage erschließen. Der erste Formant stimmt überwiegend mit dem Mundöffnungsgrad überein. Dabei bedeutet eine geringer Öffnungswinkel einen niedrigen ersten Formanten (z. B. bei /i/ oder /u/) oder einen hohen Formanten bei einer großen Öffnung, wie bei /a/. Dagegen korreliert die

Horizontallage der Zunge mit dem zweiten Formanten. Bspw. wird das /i/ vorne artikuliert und weist einen hohen zweiten Formanten auf, wohingegen das /u/ eine hintere Artikulationsstelle hat und einen niedrigeren zweiten Formanten besitzt.

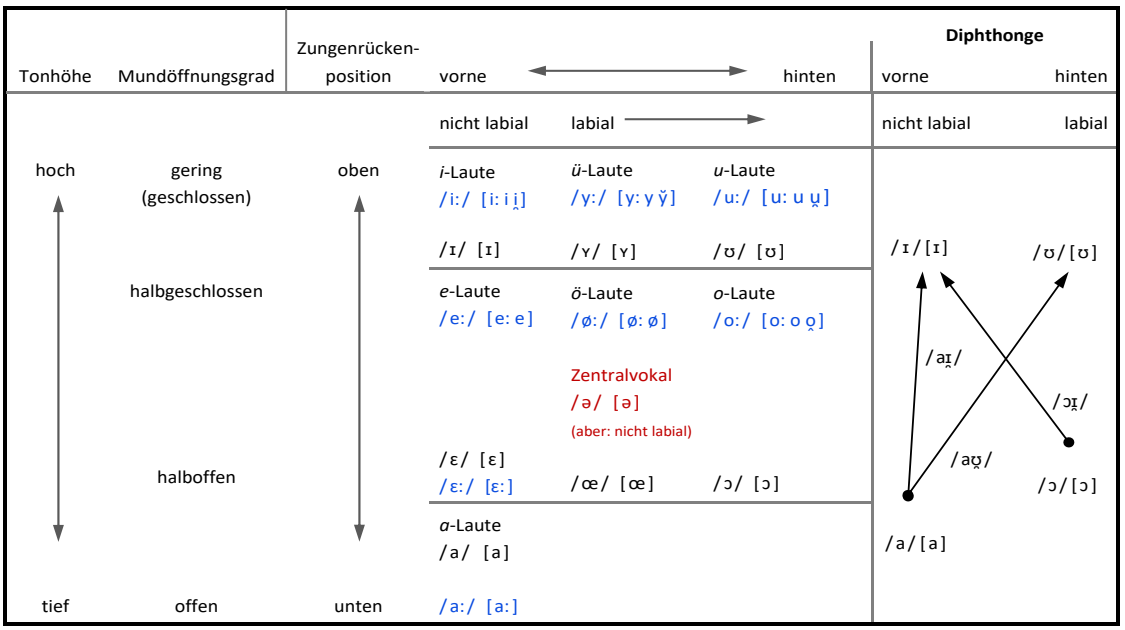


Abb. 14: Artikulatorische Eigenschaften des Vokalkernbestands des Deutschen in Relation zueinander (blau: Langvokale, schwarz: Kurzvokale), eigene Darstellung (K. Schumacher) nach [DUDEN Bd. 6 2015, S. 24-38.] und [Pompino-Marschall 2009, S. 221] entworfen

Oftmals treten diverse verschiedene Vokale hintereinander auf. Diese werden Diphthonge (Doppellaute) genannt.

> *„Phonetisch gesehen sind Diphthonge Gleitlaute, bei denen die Zunge oder die Zunge zusammen mit den Lippen eine Gleitbewegung von einer Vokalposition zu einer anderen durchführt.“*
> *[DUDEN Bd. 6 2015, S. 26]*

Im Deutschen haben nur wenige dieser Doppellaute einen Platz im Phoneminventar. Durch ihre bedeutungsunterscheidende Wirkung gehören [ai] („Reis“), [au] („Haus“) und [ɔy] („Reue“) dazu.

Akustisch gesehen sind Vokale nur von Ausprägung und Lage der Formantenfrequenzen abhängig. Die Tonhöhe eines Lautes wird im Gegensatz dazu nur durch die Grundfrequenz dirigiert. **Abb. 11** zeigt diesen Effekt.
Das mit tiefer Stimme ausgesprochene Phonem (links) weist ein sehr viel dichteres Spektrum der einzelnen Frequenzlinien auf, als das mit hoher (rechts). Es ist deutlich erkennbar, dass die Position und Ausprägung der Formanten unabhängig von der Stimmhöhe sind. [Eppinger & Herter 1993, S. 7 ff.] [Kollmeier o. D., S. 158]

Physikalische Eigenschaften
Die Erkennung von Vokalen im Sprachsignal ist leichter, da sie eine übersichtliche periodische Signalform durch die stark ausgeprägte Grundfrequenz f_g und deren

abschwächende äquidistant auftretende Harmonische[1] im Spektrogramm aufweisen. Zudem sind sie aufgrund des freien Vokaltrakts oftmals „lauter“. D. h. sie haben einen höheren Schalldruck, was im Zeitsignal gut erkennbar ist.

Wie bereits erwähnt, leisten die Formanten zur genauen Bestimmung der Vokale den Hauptbeitrag. Auch wenn bis zu fünf Formanten erkannt werden können, sind für die Differenzierung die ersten zwei Formanten ausreichend. [vgl.Eppinger & Herter 1993, S. 43 f.; Kuttruff 2004, S. 222]

Tab. 1: Frequenzbereiche der ersten beiden Formanten der Hauptvokale, [Eppinger & Herter 1993, S. 44]

Vokal	1. Formant in Hz	2. Formant in Hz
/a/	$700 - 1200$	$1000 - 1500$
/e/	$400 - 600$	$1800 - 2600$
/i/	$200 - 400$	$2000 - 3500$
/o/	$400 - 700$	$600 - 1000$
/u/	$200 - 400$	$600 - 1000$

In **Tab. 1** sind Richtwerte der Formantenfrequenzen gegeben. Sie sind in Bereiche aufgeteilt, da F_n stark streut. Die Streuungen sind vom Sprecher und der Höhe der Grundfrequenz abhängig. Es ist ersichtlich, dass jeder Vokal mit Hilfe dieser Werte bestimmt werden kann, vorausgesetzt das Sprachsignal ist deutlich genug und es tritt keine Vokalreduzierung (vgl. Kap. 2.3.1) auf. [Eppinger & Herter 1993, S. 44] Eine Abweichung aus den gegebenen Bereichen ist nach eigener Erfahrung eher ein Sonderfall und betrag zwischen $10 - 100\ \mathrm{Hz}$.

4.2.2 Konsonanten

Konsonanten werden dagegen nur durch Einengung bzw. Hinderung des Luftstroms gebildet. Manche dieser Laute werden mit Hilfe von Stimmlippenbewegungen erzeugt, d. h. sie sind stimmhaft. Ein Konsonant kann konträr zu einem Vokal nicht durch Phonem spezifische Formanten beschrieben werden. Stimmhafte Konsonanten habe ähnlich wie Vokale ausgeprägte Formanten, die jedoch nur eine Zuweisung der Phoneme zur entsprechenden Artikulationsart ermöglichen. Oftmals besitzen Konsonanten ein breites sowie kontinuierliches Spektrum (ca. $4 - 12\ \mathrm{kHz}$). Durch Konzentration ist es möglich, Artikulationsstellen bewusst wahrzunehmen. Diese werden oft zur Gruppierung der variierenden Konsonanten verwendet. In **Tab. 2** sind die verschiedenen Einteilungen aufgelistet und mit Beispielen versehen.
Entscheidend für die Konsonantenbildung ist allerdings nicht nur Ort sondern auch die Art der Luftstromhinderung, wodurch eine weitere Klassifizierung eben dieser erstrebenswert ist. Sie werden Artikulationsstellen genannt und ergeben im

[1] ganzzahlige Vielfache von f_g

Rahmen der deutschen Konsonanten fünf differente Kategorien. In **Abb. 12** ist neben einiger Vokale auch die Lautbildung von Konsonanten zu sehen.
Im Vergleich zu den Vokalen sind sie Konsonanten eine zahlreichere Phonemklasse. [Eppinger & Herter 1993, S. 9 f., S. 44] [Kuttruff 2004, S. 223] [Pfister & Kaufmann 2017, S. 15]

Tab. 2: Artikulationsstellenbezeichnung von Konsonanten, [Eppinger & Herter 1993, S. 9] [DUDEN Bd. 6 2015, S. 22 f.]

Bezeichnung	Artikulationsstelle	Beispiel
bilabial	Ober- und Unterlippe	[m], [b]
labiodental	Unterlippe und obere Schneidezähne	[f], [v]
dental	Zungenspitze und obere Schneidezähne	[s], [z]
alveolar	Zungenspitze und oberer Zahndamm (Alveolen)	[d], [n]
palatal	Zunge und harter Gaumen (Palatum)	[j], [ç]
velar	Zungenrücken und weicher Gaumen	[k], [x]
uvular	Zunge und Zäpfchen	[r], [ʁ]
glottal	Stimmritze nur teilweise geschlossen (rauschförmige Stimmbandanregung)	[h]

Bei den *Verschlusslauten* (*Explosivlaute, Plosive*) wird bei geöffneter Stimmritze der Luftstrom durch Zunge oder Gaumen gehemmt. Aufgrund des verschlossenen Nasenhöhlendurchgangs wird Druck aufgebaut, bis ein statischer Überdruck herrscht (Dauer ca. $100\ \mathrm{ms}$). Es folgt eine schlagartige Freigabe der angestauten Luft. Im Allgemeinen lassen sich Explosivlaute durch die Verschlussphase entstehende Lücke im Spektrum erkennen. Diese Aussparung ist auch im Zeitsignal erkennbar. Anschließend folgt ein kurzzeitiges vertikales Muster (spektrale Komposition), dass das Verschlusslösgeräusch wiederspiegelt.
Die Plosivlaute können in „weiche“ (stimmhafte) und „harte“ (stimmlose) untergliedert werden. Dabei besitzen die stimmhaften Verschlusslaute /b d g/ eine Stimmlippenschwingung, die meist schon früher beginnt. Stimmlose Explosivlaute wie /p t k/ können durch ihre mittlere Frequenzlage bezüglich des nachfolgenden Vokals erkannt werden. Liegt die mittlere Frequenz des Phons über $3\ \mathrm{kHz}$ so wird immer ein /t/ verstanden. Ist der zweite Formant des nachfolgenden Vokals ungefähr auf der Höhe dieser Frequenz handelt es sich um ein /k/. Aus anderen Frequenzlagen lässt sich ein /p/ erschließen.
Die *Reibelaute* (*Frikativlaute, Frikative*) /f v θ ð s z ʃ ʒ ç j x ʁ h/ entstehen mittels einer intensiven Einengung des Luftstroms durch das artikulierende Organ. Der Nasenhöhlendurchgang ist hierbei versperrt. Es entstehen Turbulenzen, die eine rauschartige Geräuschbildung nach sich ziehen.
Frikativlaute zeichnen sich durch unregelmäßig verwische Frequenzanteile im oberen Frequenzbereich über $4000\ \mathrm{Hz}$ aus. Da manche Frikative stimmhaft sind können

diese von ihrem stimmlosen Äquivalent unterscheiden werden (z. B. /s/ und /z/). Andererseits ist es auch möglich, Frikative anhand ihres verschiedenartigen Spektrums, wie bspw. bei /ʃ/ und /s/, zu differenzieren, da sich die höhere Energiedichte je nach Laut an unterschiedlichen Lagen befindet. Je weiter hinten der Frikativ artikuliert wird, desto tiefer liegt die untere Grenzfrequenz des Rauschsignals. Allgemein gilt:

> *„Je länger das der Schallquelle vorgelagerte Ansatzrohr, desto tiefer liegt die unterste Frequenz im Spektrum des entsprechenden Schallsignals". [Pompino-Marschall 2009, S. 196]*

Daraus folgt, dass das stimmhaft Phonem /ʃ/ ein tieferes Spektrum besitzt als /s/.

Nasallaute (/m n ŋ/)werden wie die Plosive durch Verschlüsse im Mundraum gebildet. Jedoch ist hier das Gaumensegel gesenkt, wodurch der Durchgang zur Nasenhöhle frei ist und der Luftstrom nur durch die Nase entweichen kann.
Sie besitzen einen stark ausgeprägten Formanten bei etwa $200\ \mathrm{Hz}$. Die weitere Formantstruktur ist nur schwach ausgeprägt. Bei nachfolgenden Vokalen ist jedoch eine abrupte spektrale Änderung des Signals erkennbar.

Im Deutschen existiert nur einer der möglichen *Seitenlaute* (*Laterale*). Die Zunge bildet beim /l/ einen Mittelsteg, an dem der Luftstrom zu beiden Seiten entweichen kann. Die Lücke ist jedoch so groß, dass der Laut nicht aufgrund einer geräuschbildenden Enge entsteht.
Der Laterallaut besitzt eine starke jedoch langsame Formantenbewegung.

Die *Vibranten* (*Schwinglaute*) werden bei verschlossenem Nasenraum durch kurz aufeinanderfolgende orale Verschlüsse hervorgerufen. Grund hierfür sind sehr elastische Artikulatoren, die einen relativ losen Verschluss an der entsprechenden Artikulationsstelle bilden. Dieser wird in der nächsten Phase durch den ausströmenden Luftstrom gesprengt. Aufgrund der stabilen Verengung durch den jeweiligen angespannten Artikulator wird der Bernoulli-Effekt[2] ausgelöst, was zu einem erneuten Verschluss führt. Diese Abfolge wird zwei- bis dreimal durchlaufen, wobei ein Rollen an Zungenspitze oder Zäpfchen entsteht.
Im Spektrum sind die Vibranten nicht oder nur kaum ersichtlich. Dagegen können sie aufgrund einer niederfrequenten Amplitudenmodulation im Zeitsignal erkannt werden.[DUDEN Bd. 6 2015, S. 26 f.] [Eppinger & Herter 1993, S. 10, S. 45 f.] [Kuttruff 2004, S. 223 f., S. 232] [Pompino-Marschall 2009, S. 130, S.184 ff.] [Terhardt 1998, S. 198]

[2] Eine Verengung der Durchflussöffnung führt zur Erhöhung der dort auftretenden Fließgeschwindigkeit des Luftstromes. Es entsteht ein Unterdruck, wodurch die Bernoulli-Kräfte eine Sogwirkung senkrecht zur Fließrichtung bewirken und einen erneuten Verschluss verursachen. [Pompino-Marschall 2009, S. 32-34]

4.2.3 Merkmale eines Sprachsignals

Normales bzw. unangestrengtes Sprechen liegt bei etwa $65\,\mathrm{dB}$. Der natürliche Sprachfluss weißt – hauptsächlich aufgrund von Vokalen und Plosivlauten – einschneidende Merkmale auf, die eine zeitlichen Untergliederung herbeiführen. Vokale werden im Gegensatz zu Konsonanten länger gehalten und weisen eine deutlich höhere spektrale Energiedichte auf. Plosivlaute sind durch die jeweilige präplosive Pause erkennbar (vgl. Kap. 4.2.2). Diese sind vergleichsweise kürzer als Atem- oder Sprechpausen. Die Anzahl der gesprochenen Laute in der Sekunde (*Lautfrequenz*) beträgt $8 - 15\ \frac{1}{s}$ und ist stark vom Sprecher abhängig. Die mittlere Stimmbandschwingungsfrequenz (Grundfrequenz oder auch Grundton) beträgt bei Männern $120\ \mathrm{Hz}$ und bei Frauen $240\ \mathrm{Hz}$. Sie variiert bei jedem Sprecher um ca. eine Oktave und kann grundsätzlich zwischen $70 \ldots 500\ \mathrm{Hz}$ betragen.

Der Frequenzbereich eines Sprachsignals ist nach unten durch die Stimmlippen-Oszillationsfrequenz auf $70 \ldots 100\ \mathrm{Hz}$ beschränkt. Eine absolute Obergrenze ist nicht vorhanden. Eine bedeutende Rolle spielen jedoch nur Frequenzen bis etwa $10\ \mathrm{kHz}$. In diesem Bereich (auch bei noch höheren Frequenzen) liegen nur die Frikativ- und Plosivlaute.
Ein Sprachsignal weist nie exakt periodische Verläufe auf. Es entstehen aufgrund der ausströmenden Luft und der quasi-periodischen Anregung des Luftstroms der Stimmlippen Überlagerungen, die sich in der Zeitsignal-Darstellung wiederspiegeln. [Terhardt 1998, S. 198 ff.] [Pfister & Kaufmann 2017, S. 45]

Spektrogramm (3D-Darstellung, Sonagramm)
Im Gegensatz zum Frequenzspektrum[3] kann beim *Spektrogramm* der Verlauf über die Zeit dargestellt werden und erleichtert somit die Erkennung der Laut- bzw. Sprachmerkmale. Zur Ermittlung eines Spektrogramms wird das aufgenommene Sprachsignal in kleine gleichmäßige Abschnitte unterteilt. Von jedem dieser Abschnitte wird nun das Betrags- und Leistungsdichtespektrum mittels Fouriertransformation kalkuliert. Bei diesem dreidimensionalen Diagramm ist der Schalldruckpegel[4] farblich kodiert dargestellt und kann mit Hilfe einer Farbskala abgeschätzt werden. Auf der vertikalen Achse (Ordinate) ist die Frequenz (meist logarithmisch) und auf der horizontalen Achse (Abszisse) die Zeit angegeben.

Das *Visible Speech Diagramm* war der Ursprung dieser Darstellungsvariante und gibt die Intensität der Frequenzbänder in variierender Schwärzungsintensivität wieder. Dieses Verfahren wurde bereits nach dem zweiten Weltkrieg entwickelt und lieferte damals das Ergebnis auf einem Registrierstreifen, dass die oben genannten Dimensionen besaß. Zudem haben zu dieser Zeit praktische Versuche ergeben, dass das Lesen dieser „sichtbaren Sprache“ erlernbar ist. [Eppinger & Herter 1993, S. 40 f.] [Meyer & Neumann 1967, S. 233] [Pfister & Kaufmann 2017, S. 48]

[3] Darstellung der Energie über die Frequenz
[4] Größe zur Beschreibung der Stärke eines Schallereignisses

4.3 Analyse

Die Probanden (vgl. Kap. 3) werden wie folgt adressiert:

- männliche Sprecher:
  mS1 ($f_g = 130$ Hz)
  mS2 ($f_g = 110$ Hz)
- weibliche Sprecher:
  wS1 ($f_g = 220$ Hz)
  wS2 ($f_g = 210$ Hz)

Die weiblichen Sprecher haben eine tiefere Stimme als der Durchschnitt ihres Geschlechts (vgl. Kap. 4.2.3). Der Sprecher mS1 liegt mit 10 Hz darüber und mS2 darunter. Die Werte wurden über mehrere Aufnahmen er- bzw. gemittelt. Die Spektrogramm-Erstellung in *ArtemiS* erfolgte im Bereich von 70 ... 19.000 Hz, sodass die Frequenzen weitgenug abgedeckt sind. Die Skala der Spektrogramme sind grundsätzlich logarithmisch unterteilt.

Zur leichteren Unterscheidung wurden nur farbliche Spektrogramme für die Analyse der Sprachsignale verwendet. Eine richtige 3D - Darstellung ist erst bei der Änderung des erzeugten Diagrammes von *spectrogram* auf *3D* möglich. Diese Ansicht unterstützt in manchen Situationen die Erkennung der Formanten. Jedoch ist sie für weiterführende Betrachtungen eher ungeeignet.

4.3.1 Analyse einzelner Laute

Das mit Hilfe eines Mikrofons aufgenommene Sprachsignal beinhaltet im zeitlichen Verlauf nur geringfügig Informationen über die geäußerten Phonemfolgen. [Kollmeier o. D., S. 156]. Durch eine vergrößerte Zeitachse ist es möglich, die Periodizität, also die Stimmhaftigkeit, eines Lautes festzustellen. Für die Lautübergänge ist ein Spektrogramm besser geeignet. Die von Kollmeier (o. D.) erkannte Eigenart des Zeitsignals wurde ebenfalls bei den eigenen Analysen festgestellt. Es ist bspw. möglich den periodischen Verlauf eines stimmhaften Lautes bei isolierter Einsprache sowie in einem Wort gut zu erkennen, vorausgesetzt der Proband hält den Laut lange genug, sodass das Signal des Phons lang genug für eine Auswertung ist. Zudem sind sowohl Amplitude als auch Art der periodischen Verläufe bei demselben Sprecher jedes Mal unterschiedlich. **Abb. 15** zeigt diesen Effekt.

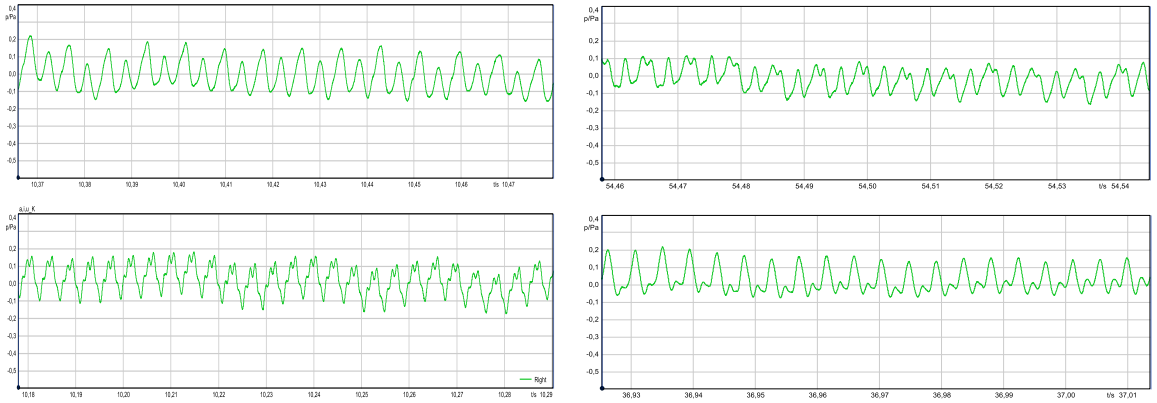


Abb. 15: Vokal /u/ bei ben Sprechern mS2 (oben) und wS2(unten) als einzeln eingesprochener Laut (links) und in dem Wort Student (rechts), ArtemiS

Es ist das Sprachsignal eines männlichen Sprechers (erste Reihe) und eines weiblichen Sprechers (zweite Reihe) zu sehn. Rechts ist der Laut mit dem Wort „Student“ und links einzeln eingesprochen worden. Auffällig ist hierbei, dass die Periodizitäten beim selben Sprecher gleichbleiben.

Die Darstellung der Konsonanten ist nur im Spektrogramm sinnvoll, da die meisten ein rauschartiges Signal besitzen und somit keine eindeutige Erkennung ermöglichen.
In **Abb. 16** sind Beispiele für Frikativ-, Explosiv-, Lateral- und Nasallaute gegeben. Wie in Kap. 4.2.2 erwähnt, ist das Spektrum über die Zeit von /f/ breitbandiger und etwas niederfrequenter als beim /s/, da der Luftstrom des Lautes weiter hinten eingeengt wird. Das /z/ wird im Deutschen als das „stimmhafte s“ bezeichnet und weist ein ähnliches Spektrum wie das /s/ auf, mit dem Unterschied, dass es ein Mitschwingen der Stimmbänder beinhaltet.

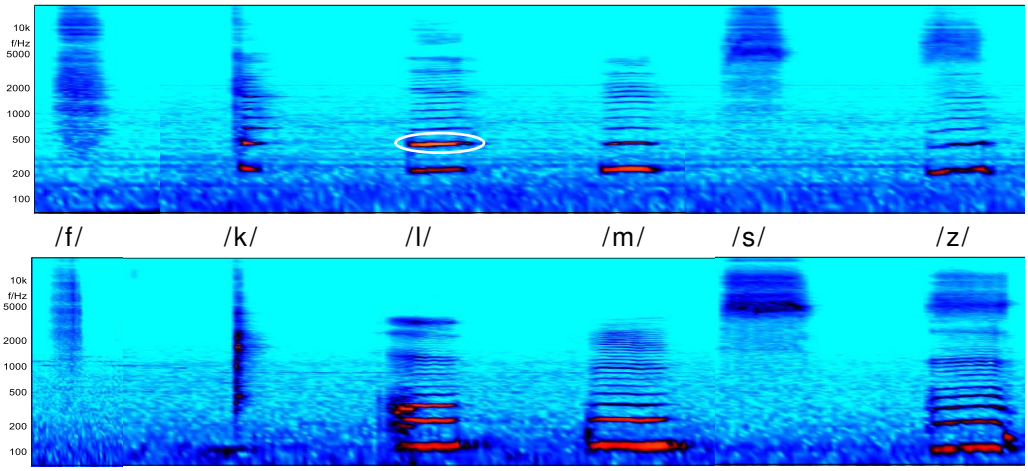


Abb. 16: Spektrogrammdarstellung einiger Konsonanten des Sprechers wS1 (oben) und mS1 (unten), ArtemiS

Der Verschlusslaut /k/ hat zuerst eine breite spektrale Komposition und besitzt bei Phonemketten üblicherweise keinen Grundton. Der weibliche Sprecher wS1 artikulierte demnach mit Glottisschwingung, wie es beim Aufsagen des ABCs üblich ist. Die Lücke im Spektrum ist aufgrund der isolierten Einsprache nicht gegeben.
Nasallaute wie das /m/ besitzen bei ungefähr $200\ \mathrm{Hz}$ einen starken Formanten, der auch hier zu sehen ist. Dieser ist bei männlichen Sprechern leichter zu erkennen, da f_g bei den weiblichen Probanden sehr nah an dieser Frequenz liegt.
Während der Analysen ist besonders der Frequenzbereich um $400\ \mathrm{Hz}$ bei dem Laterallaut /l/ ins Auge gefallen, siehe obere Reihe **Abb. 16**. Dieser Bereich streut gelegentlich und ist in manchen Fällen nicht allzu deutlich, dennoch zeichnete sich bei allen vier Sprechern und verschiedenen Wortanalysen ein deutliches Muster ab.

4.3.2 Analyse von Wörtern und Wortketten

Werden nun die Laute in Verbindung gesetzt entstehen diverse Übergänge, die bei schneller Sprache schwer zu erkennen sind. Deshalb wurden für die bessere Verständlichkeit langsame Sprachausschnitte verwendet. In **Abb. 17** ist sowohl das Zeitsignal (oben) als auch das Spektrogramm (unten) des Wortes „Dorfplatz“

([ˈdɔʁfplats]) der Sprecherin wS2 gegeben. Die Formanten den Phoneme [ɔ] und [a] sind mit einem weißen Strich gekennzeichnet und betragen:

- [ɔ]: $\mathrm{F}_1 = 440\ \mathrm{Hz}$, $\mathrm{F}_2 = 650\ \mathrm{Hz}$
- [a]: $\mathrm{F}_1 = 750\ \mathrm{Hz}$, $\mathrm{F}_2 = 1200\ \mathrm{Hz}$.

Sie stimmen mit der in **Tab. 1** aufgezeigten Richtwerten überein.
Im Zeitsignal ist erkennbar, dass die Amplituden der Vokale gegenüber den Konsonanten deutlich höher sind, sowie eine Lücke vor [p] und [t] existiert. Der weiche Explosivlaut [d] hat einen schlagartigen Anfang und ein Mitschwingen der Stimmbänder (Grundton und Harmonische). Dagegen haben die „harten“ Verschlusslaute nur eine abrupte Entstehung.

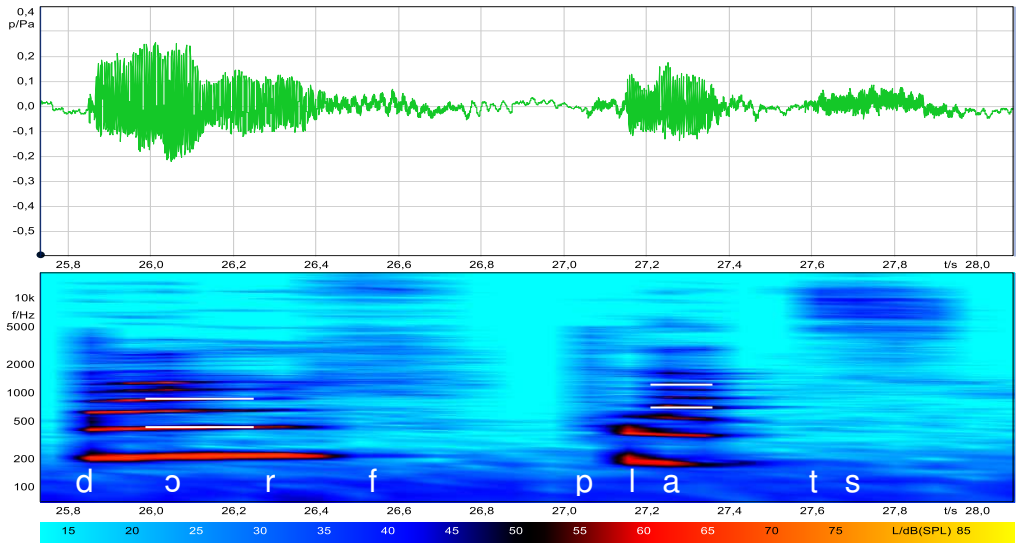


Abb. 17: Wort "Dorfplatz" von wS2, ArtemiS

Das Phonem [t] hat eine mittlere Frequenz über $3\ \mathrm{kHz}$ und korreliert mit der aus Literaturrecherchen entnommenen in Kap. 4.2.2 dargelegten Erkenntnis. Ein Vergleich mit der mittleren Frequenz zur Erschließung eines [p] ist ausgeschlossen, da [l] kein Vokal ist und somit nicht die vorausgesetzte Bedingung erfüllt. Es lassen sich dadurch keine weiteren Rückschlüsse auf das einzelne Phonem schließen.

Das [r] ist in beiden Grafiken - selbst mit deutlicher Vergrößerung - nicht zu erkennen. Das Phonem [f] weist ein breites Spektrum auf. Ebenso könnte es sich in dem Fall um die „ch“ – Laute ([ç] bzw. [x]) handeln. Ein genauer Unterschied wäre hier nicht durchführbar. Wie zuvor erwähnt, besitzt das [l] einen ausgeprägteren Formantenbereich bei $400\ \mathrm{Hz}$ und das [s] hat eine vordere Artikulationsstelle, wodurch es ein hohes Frequenzspektrum über $4\ \mathrm{kHz}$ erreicht wird.
Die Analyse wurden auch mit stimmlos ausgesprochenen Wörtern durchgeführt. Der Hauptunterschied liegt in der Entwirrung der Lautfolgen, da nun keine Grundfrequenz oder deren Harmonische erkenntlich sind. Allerdings erreichen die Formanten der Vokale eine verhältnismäßig starke Ausprägung. Besonders das /a/ ließ sich noch gut erkennen, da die Formantenbereiche ineinander übergehen und so um $1\ \mathrm{kHz}$ spektrale Anteile mit einem hohen Energiegehalt aufweisen.

Abb. 18 gibt den Verlauf von fließender Sprache (normale Sprechgeschwindigkeit) wieder. Für eine flüssige Lesbarkeit ist in der Abbildung die tatsächliche schriftliche

Fassung und nicht die Lautschrift angegeben. Es enthält die typischen Merkmale wie ein höherliegender ausgeprägter Formantenbereich bei [s] (z. B. des) im Vergleich zum stimmhaften [ʃ] (Student). Ebenfalls ist ersichtlich, welcher Laut einen Plosivlaut darstellt (vgl. bei $1{,}5\ s$ und $6{,}5\ s$). Die Lautgrenzen sind jedoch so ineinander übergreifend, dass es ohne abhören des Sprachsignals oftmals schwierig ist,

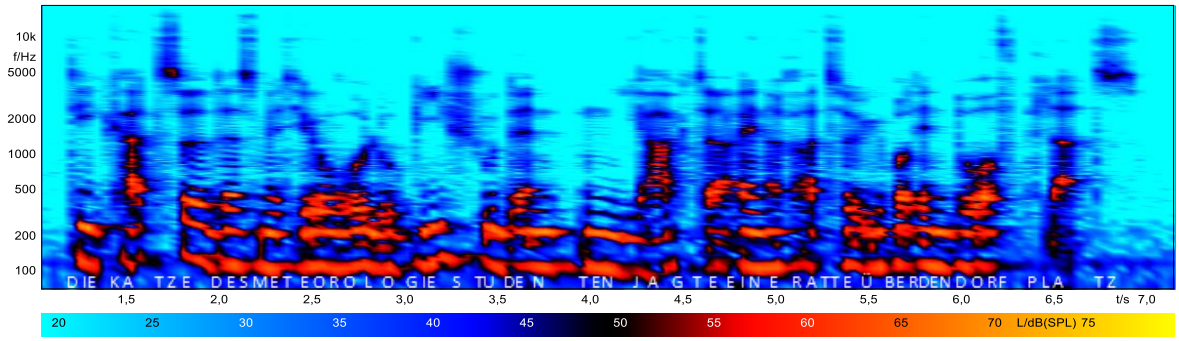


Abb. 18: Spektrogramm von fließender Sprache (mS2), ArtemiS

die Laute voneinander zu trennen.
Der Satz enthält 57 Phone[5] und beträgt beim Sprecher mS2 ca. $6{,}5\ s$. Die Lautfrequenz dieses Sprechers beträgt somit $8{,}8\ \frac{1}{s}$. Die gesprochenen Laute pro Sekunde der anderen Sprecher liegen bei: mS1: $13{,}6\ \frac{1}{s}$; wS1: $10{,}4\ \frac{1}{s}$ und wS2: $12{,}1\ \frac{1}{s}$.
Alle Werte entsprechen der durchschnittlichen Sprechgeschwindigkeit (siehe Kap. 4.2.3).

Aufgrund weiterführenden Denkens entwickelte sich die Idee, ob es innerhalb dieses Projektes möglich sei, ein Wort nur durch Zeitsignal und Spektrogramm zu erraten. Vier verschiedene Wörter wurden eingesprochen, jedoch konnte - trotz über $85\ \%$ erkannter Laute - nie das gewünschtes Ergebnis erzielt werden.
Die Wörter betrugen eine unterschiedliche Anzahl an Phonemen. Die „langen“ Worte konnten aufgrund der starken Beeinflussung benachbarter Laute (siehe Kap. 2.3.1) nicht entschlüsselt werden. Eine Dechiffrierung der „kurzen“ Phoneketten war aufgrund einzelner nicht erkannter Phoneme (z. B. [r]) unerreichbar. Der Hauptgrund für diesen erfolglosen Versuch ist höchstwahrscheinlich ein Mangel an Erfahrung bzw. einer zu kurzen Übungszeit für den Leseerwerb der „sichtbaren Sprache“.

4.3.3 Resümee

Im Allgemeinen unterscheiden sich die Phoneme bei den unterschiedlichen Sprecher nur geringfügig. Die Formanten der Vokale sind bei männlichen Sprechern tiefer und oft leichter erkennbar. Konsonanten sind gering vom Sprechertyp abhängig. Nasal-, Plosiv- und Laterallaute wiesen keinen Unterschied auf. Die einzige hauptauffällige Besonderheit erschien bei den Frikativen. Die Sprecherin wS2 besitzt - im Vergleich zu den anderen Sprechern - bei dem Phon [s] einen um etwa $2 \ldots 3\ \mathrm{kHz}$ nach oben versetzten Frequenzbereich.

[5] di: ˈkatsə dɛs meteoʁoloˈgi: ʃtuˈdɛntn̩ ˈja:ktə ˈaɪ̯nə ˈʁatə ˈy:bɐ de:n ˈdɔʁfˌplats

5 Psychoakustische Analyse

5.1 Vorwort

Ziel dieser Untersuchung ist es, herauszufinden ob und wie sich psychoakustische Parameter zur Analyse von Sprache eignen. Die dazu ausgewählten Kenngrößen sind:

- Lautheit
- Tonhaltigkeit
- Schärfe
- Schwankungsstärke
- Rauigkeit

Die Auswahl der Parameter ist eher zufällig getroffen wurden. Es gibt weitere Kenngrößen, die für eine Betrachtung nützlich sein könnten, doch aus Zeitgründen nicht weiterverfolgt wurden.

5.2 Psychoakustische Parameter

Lautheit

Die Tendenzen der Lautheit sind recht einheitlich (**Abb. 19**). Jeder Sprecher entscheidet sich zu Beginn einer Äußerung für eine bestimmte Lautheit. Die darauffolgenden Äußerungen orientieren sich von der Lautstärke her stark an dem zuvor gesagten. Dadurch ist es möglich die Lautheit selbst über längere Gespräche relativ konstant zu halten.

Beim Reden werden natürlich verschiedene Worte und Laute unterschiedlich betont. Die Abweichungen sind dennoch gering. Die höchste Lautheit besitzt meist ungefähr den doppelten Wert der geringsten.

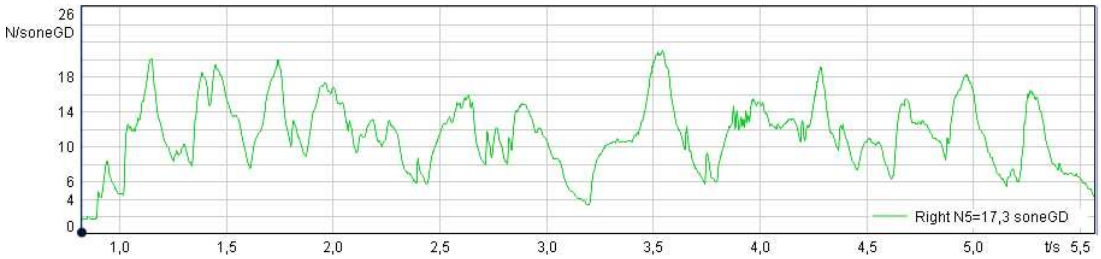


Abb. 19: Lautheit eines gesprochenen Satzes (wS2), ArtemiS

Tonhaltigkeit

Dieser Parameter reagiert immer und ausschließlich auf stimmenhafte Anregungen. Die Grundfrequenz und deren Harmonischen sind somit im Klangbild am deutlichsten wahrnehmbar. Je stärker deren Anregung und je schwächer ausgebildet die Formanten, desto markanter das Signal. Dies bewahrheitete sich bei allen Sprechern.

In der rechten Grafik in **Abb. 20** ist eine FFT dargestellt, welche von der linken Tonhaltigkeit ergänzt wird. Das Sprachsignal beinhaltet die 6 Laute „l, m, n, o, p und k“. Auffällig ist, wie bei den stimmenhaften Lauten „l, m, n, o“ die Tonhaltigkeit ausschlägt und bei den stimmlosen Plosiven „p, k“ keine wirkliche Signaländerung zu verzeichnen ist.

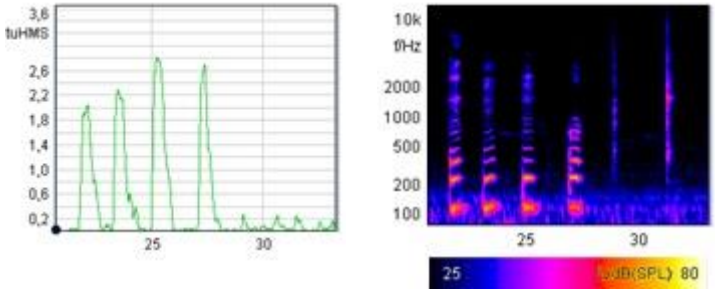
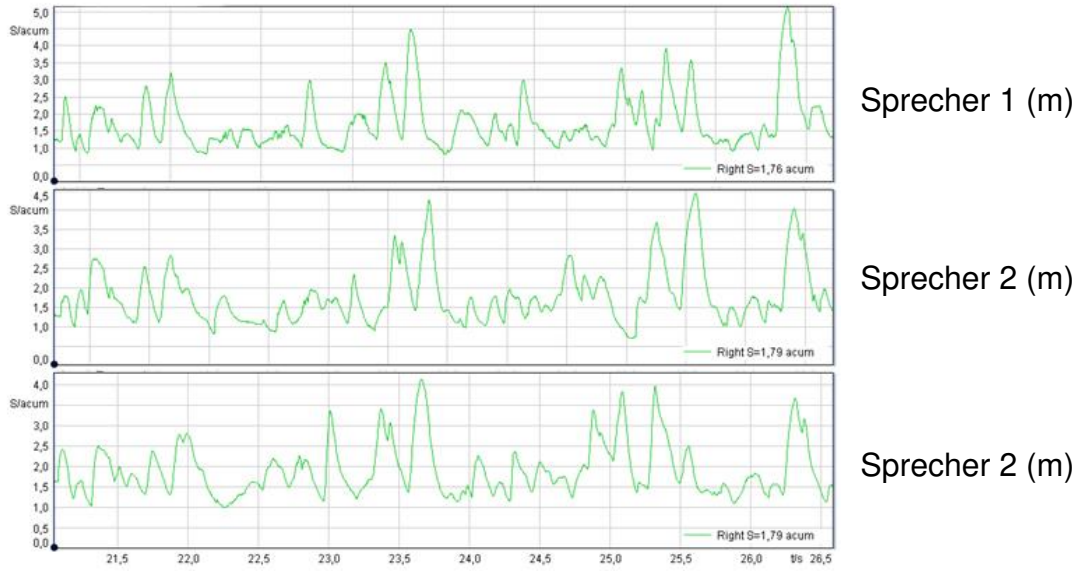


Abb. 20: Tonhaltigkeit einzelner Laute

Schärfe

Dieser Parameter zeichnet sich durch das zuverlässige Erkennen von Frikativen aus. Auffällig bei der Kenngröße ist, wie identisch sich die Signale aller Probanden im Vergleich sehen.



Sprecher 1 (m)

Sprecher 2 (m)

Sprecher 2 (m)

Abb. 21: Schärfe eines gesprochenen Satzes, Vergleich; ArtemiS

Das Signal von Sprecher 4 (m) ist nicht enthalten, da das Timing nicht stimmte und die Darstellung dadurch ablenkend wirkte.

Schwankungsstärke

Dieser Parameter hat sich als unnütz zur Untersuchung von Sprache erwiesen (vgl. **Abb. 22**). Um zuverlässige Signale zu erzeugen benötigt man Klänge, die mindestens 4 Sekunden lang gehalten werden. Da dies nicht der Fall in Gesprächen ist, können deshalb keine aufschlussreichen Ergebnisse generiert werden.

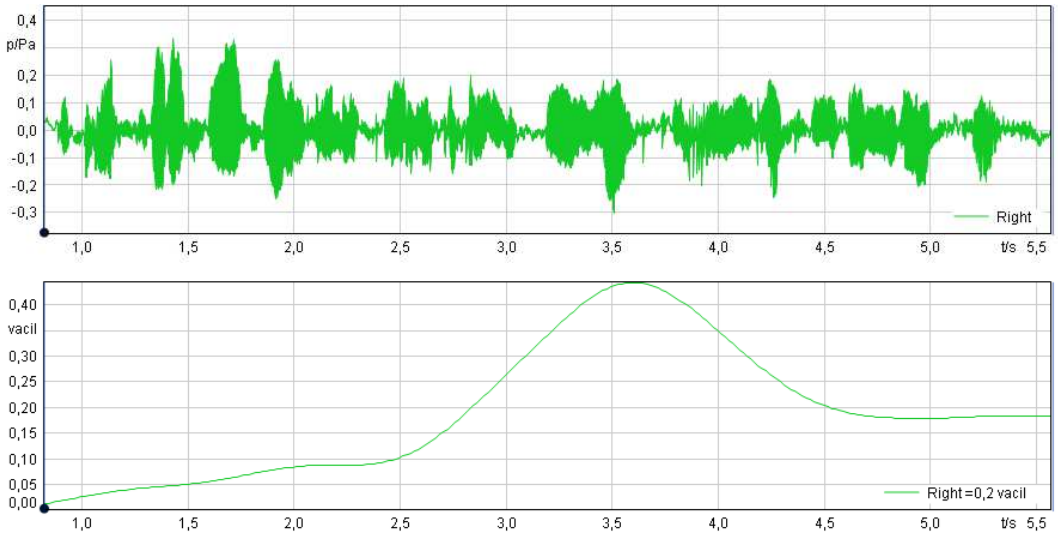


Abb. 22: Schwankungsstärke eines gesprochenen Satzes, mit Mikrofonsignal

Rauhigkeit

Ausschläge treten hauptsächlich bei stimmenhaften Anregungen auf, die unsauber gehalten werden. Im Frequenzspektrum ist dies bei genauer Betrachtung leicht erkenntlich. Sobald die Grundfrequenz und deren Harmonische anfangen eine „Zitterbewegung“ durchzuführen, steigt die Rauigkeit rapide an.

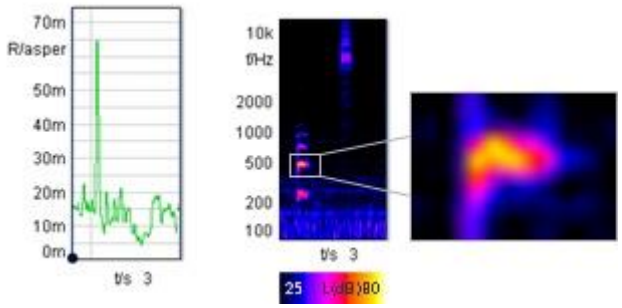


Abb. 23: Rauigkeit einzelner Laute; rechts: FFT mit Vergrößerung der ersten Harmonischen, links: entsprechende Rauigkeit zur FFT

5.3 Resultat

Einige Parameter liefern vielversprechende Ergebnisse. Das konstant halten der Lautheit, das Überprüfen der Stärke der Tonhaltigkeit bei stimmhaften Lauten sowie das Erkennen von Frikativen durch die Schärfe und das saubere halten von stimmhaften Tönen durch die Rauigkeit könnten genutzt werden, um Sprachkurse zu gestalten. Mit der Schwankung konnten zwar in diesen Versuchen keine Ergebnisse erzielt werden, doch vorstellbar wären z.B. Anwendungen im Gesangsunterricht.

Grundsätzlich lassen sich mit den Kenngrößen Klangqualitäten überprüfen, die anderen falls schwer erkenntlich wären. Die daraus resultierenden Erkenntnisse sind auf die gesprochene Sprache anwendbar und somit vielseitig einsetzbar.

6 Zusammenfassung

Die Handhabung von Sprache ist ein großes komplexes Konstrukt. Es fängt damit an, dass Menschen seit der Geburt zum Sprechen aufgefordert werden. Es ist ein wichtiger Bestandteil des Heranwachsens und es dauert bis zum vierten Lebensjahr um komplexe Sachverhalte verständlich darzulegen. Eine fehlerfreie Sprache ist sogar erst ab Grundschulalter möglich. Die Entstehung dieser Sprachlaute erfolgt durch Lunge, Kehlkopf und Stimmritze. Der Vokaltrakt stellt die schallmodifizierende Komponente dar und bestimmt um welchen Laut es sich handelt. Es können unendlich viele Lautvariationen auftreten. Zur Verständigung untereinander wird jedoch nur eine kleine Menge genutzt. Diese können in stimmhaft und stimmlos sowie Vokale und Konsonanten kategorisiert werden. Die fließende Sprache ist stark von den vorherigen und nachfolgenden Phonen oder Wörtern abhängig. Dieser Effekt wird unter dem Wort Koartikulation verstanden. Lautliche Eigenschaften wie Tempo oder Rhythmus werden unter der Bezeichnung Prosodie zusammengefasst. Beide Überbegriffe bilden die Hauptunterscheidungsmerkmale des Gesagten. Es können bspw. bei ungenauer oder zu kurzer Vokalaussprache eine Vokalreduzierung auftreten. Die Untersuchung solcher Signale kann dann nur schwer erfolgen auch wenn das Gehör alles verstanden hat. Aufgrund der technischen Evolution wurde die Spracherzeugung in eine idealisierte Modellvorstellung umgewandelt um eine künstliche Nachbildung erzeugen zu können.

Die linguistischen Merkmale einer Sprache sind sehr weitreichend. Diese wurden hier auf das Wesentliche reduziert. Dazu gehören unteranderem die Phoneme, welche als kleinste bedeutungsunterscheidende Einheiten gelten. Allophone dagegen sind die unterschiedlichsten Artikulationsvarianten dieser.
Wie zuvor erwähnt, können die Sprachlaute in die Lautklassen Vokale und Konsonanten unterteilt werden. Die Vokale zeichnen sich im Zeitsignal durch einen hohen Schalldruck aus und können durch eine Spektrogramm-Darstellung mit Hilfe der Formanten spezifiziert werden, da sie alle stimmhaft sind. Konsonanten dagegen haben oftmals ein unregelmäßiges Spektrum und können vorerst nur in die verschiedenen Artikulationsarten unterteilt werden. Eine weitere Klassifizierung ist nicht in allen Fällen möglich. Die eigenen Aufnahmen spiegeln häufig die aus der Literatur entnommenen Erkenntnisse wieder, allerdings konnten nicht alle der physikalischen Eigenschaften erkannt werden.
Analysen diverser psychoakustischer Größen ergaben, dass sich mit Hilfe von Komponenten wie Rauigkeit oder Tonhaltigkeit verwertbare Ergebnisse erzielen lassen.

Auf den Punkt gebracht: Das Gehör ist unübertroffen wenn es darum geht gesprochene Laute oder Lautverbindungen zu erkennen bzw. zu verstehen.

VI Literaturverzeichnis

[DUDEN | Aussprache (o. D.)]

DUDEN | AUSSPRACHE (o.D.). [Lautschrift nach IPA]. Abgerufen von https://www.duden.de/hilfe/aussprache (2. März 2019).

[DUDEN Bd. 6 2015]

DUDEN BD. 6 (2015). *Das Aussprachewörterbuch: Betonung und Aussprache von über 132.000 Wörtern und Namen* (7. Aufl.). Berlin: Bibliographisches Institut.

[Eppinger & Herter1993]

EPPINGER, B., & HERTER, E. (1993). *Sprachverarbeitung*. München Wien: Carl Hanser.

[Fodor o. D.]

FODOR, T. (o. D.). [Erlernung von Sprache]. Abgerufen von http://www.szulo-klapja.hu/gyermek-fejlodese/836/hogyan-segithetjuk-a-gyermek-beszedfejlodeset-utmutato-0-7-eves-korig.html (19. Februar 2019).

[Kollmeier o. D.]

KOLLMEIER, B. (o. D.). *Audiologie* [Vorlesungsscript]. Abgerufen von http://medi.uni-oldenburg.de/download/docs/lehre/kollm_audiologie/ (27. Februar 2019).

[Kuttruff 2004]

KUTTRUFF, H. (2004). *Akustik: eine Einführung*. Stuttgart: Hirzel.

[Meyer & Neumann 1967]

MEYER, E., & NEUMANN, E. (1967). *Physikalische und Technische Akustik: Eine Einführung mit zahlreichen Versuchsbeschreibungen.* Braunschweig: Friedr. Vieweg & Sohn.

[Pfister & Kaufmann 2017]

PFISTER, B., & KAUFMANN, T. (2017). *Sprachverarbeitung: Grundlagen und Methoden der Sprachsynthese und Spracherkennung*. Berlin Heidelberg: Springer.

[Pompino-Marschall 2009]

POMPINO-MARSCHALL, B. (2009*). Einführung in die Phonetik* (3. Aufl.). Berlin: Walter de Gruyter.

[Terhardt 1998]

TERHARDT, E. (1998). *Akustische Kommunikation: Grundlagen mit Hörbeispielen*. Berlin Heidelberg: Springer.

VII Anlagenverzeichnis

Anlagen, Teil 1

Einstellungsparameter FFT vs. Time und Ergänzung

Es wurden mehrere Versuche mit unterschiedlichen Einstellung in ArtemiS gearbeitet, allerdings haben die Standardeinstellungen (siehe **Abb. 24**) das beste Ergebnis erzieht. Außerdem ist noch eine Abbildung des Wortes „Bestand“ (**Abb. 25**) in 3D zur Vervollständigung dargestellt. Sie befindet sich im Anhang, da diese für die Analyse nicht genutzt wurde.

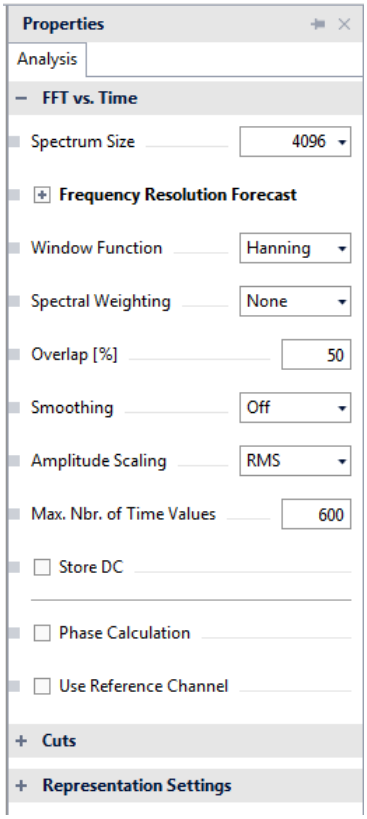


Abb. 24: Standardeinstellungen der FFT vs. Time in ArtemiS

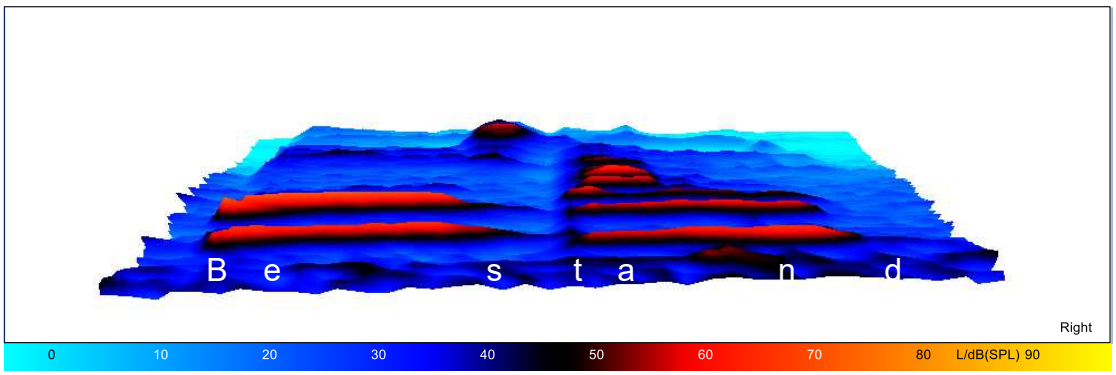


Abb. 25: 3D-Darstellung des Wortes Bestand (wS2), ArtemiS

Anlagen, Teil 2

Phoneminventar und verwendete Zeichen

Im Folgenden ist der Hauptkern des deutschen Phoneminventars angegeben. Zudem ist eine Tabelle abgebildet, die Zeichenbedeutungen angibt. Beide Darstellungen sollen zu einem besseren Verständnis der verwendeten Zeichen verhelfen.

Tab. 3: Zeichenbedeutungen, [DUDEN Bd. 6 2015, S. 12 f.] [DUDEN | Aussprache (o. D.)]

Zeichnen	Bedeutung
ː	Längenzeichen: Betonung der Länge des vorangehenden Lautes
ˈ	Hauptakzent: steht vor hauptbetonender Lautfolge
ˌ	Nebenakzent: steht vor nebenbetonender Lautfolge
◌̯ ◌̆	Kennzeichnung unsilbischer Vokalbildung
[ ]	phonetische Lautschrift
/ /	phonemische Lautschrift

Tab. 4: Lautschrift nach IPA (nicht vollständig), [DUDEN | Aussprache (o. D.)]

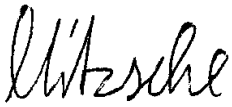
Phonem	Beispiel	Aussprache
Vokale		
/**a**/	hat	hat
/**a**ː/	Bahn	baːn
/**ɐ**/	Ober	'oːbɐ
/**e**/	Methan	me'taːn
/**e**ː/	Beet	beːt
/**ɛ**/	hätte	'hɛtə
/**ə**/	halte	'haltə
/**i**/	vital	vi'taːl
/**i**ː/	viel	fiːl
/**ɪ**/	Birke	'bɪrkə
/**o**/	Moral	mo'raːl
/**ø**/	Ökonom	øko'noːm
/**ɔ**/	Post	pɔst
/**u**/	kulant	ku'lant

/**ʊ**/	Pult	pʊlt
/**y**/	Physik	fy'ziːk
/**yː**/	Rübe	'ryːbə
/**ʏ**/	füllen	'fʏlən
Diphthonge		
/**a͡ɪ**/	weit	va͡ɪt
/**a͡ʊ**/	Haut	ha͡ʊt
/**ɔ͡ɪ**/	Heu	hɔ͡ɪ
Konsonanten		
/**b**/	Ball	bal
/**ç**/	ich	ɪç
/**d**/	dann	dan
/**f**/	viel	fiːl
/**g**/	gut	guːt
/**h**/	hat	hat
/**j**/	ja	jaː
/**k**/	kalt	kalt
/**l**/	Last	last
/**m**/	Mast	mast
/**n**/	Naht	naːt
/**n̩**/	baden	'baːdn̩
/**ŋ**/	lang	laŋ
/**p**/	Pakt	pakt
/**r**/	Rast	rast
/**ʁ**/	Dorf	dɔʁf
/**s**/	Wasser	'vasɐ
/**ʃ**/	Schal	ʃaːl
/**t**/	Tal	taːl
/**v**/	was	vas
/**x**/	Bach	bax
/**z**/	Hase	'haːzə
/**ʒ**/	Genie	ʒe'niː
/**ð**/	on the rocks	ɔn ðə'rɔks (engl.)
/**θ**/	Synthesizer	'sɪnθɪsa͡ɪzə (engl.)

Selbstständigkeitserklärung

Hiermit erklären wir, dass die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt wurde.
Stellen, die wörtlich oder sinngemäß aus Quellen entnommen wurden, sind als solche kenntlich gemacht.
Diese Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt.

Mittweida, den 11. März 2019



Robert Nitzsche

Kathrin Schumacher

Péter Tóth