universität wien

MAGISTERARBEIT / MASTER'S THESIS

Titel der Magisterarbeit / Title of the Master's Thesis

„Stimme als akustisches Mittel in der Fernsehwerbung.

Kompetenz- und Aufmerksamkeitssignale."

verfasst von / submitted by

Dorothea Stepan Bakk. phil.

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magistra der Philosophie (Mag. phil.)

Wien, 2017 / Vienna 2017

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 841
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Publizistik und Kommunikationswissenschaft
Betreut von / Supervisor:	Univ.-Prof. Dr. Jürgen Grimm

DANKSAGUNG

Mein besonderer Dank gilt folgenden Personen: Herrn Univ.-Prof. Dr. Jürgen Grimm für seine Geduld und sein Verständnis dafür, dass das Verfassen einer wissenschaftlichen Arbeit neben der vollen Berufstätigkeit entsprechend lange dauern kann. Herrn Univ.-Prof. Dr. Dr. Michael G. Schimek für wertvolle Diskussionen und konstruktive Kritik. Meinem Lebenspartner für sein Verständnis dafür, dass in den letzten Jahren ein Großteil meiner Urlaubstage dem Verfassen der vorliegenden Arbeit gewidmet war.

INHALTSVERZEICHNIS

ABBILDUNGS- UND TABELLENVERZEICHNIS

1. EINLEITUNG

Allgemein gesprochen, geht es in dieser Arbeit um den Einsatz von Stimmen in audiovisuellen Medien. Konkret geht es um die Frage, ob kommunikative Ziele der Werbung als prosodische Merkmale von Sprechstimmen identifiziert werden können.

Die Stimme als Überbringerin (Medium) einer gesprochenen Botschaft beeinflusst dieselbe entscheidend. Wird zum Beispiel eine Botschaft von einer hohen Frauenstimme gesprochen hat dies einen anderen Klangcharakter als bei einer tiefen Männerstimme und führt dadurch auch zu einem anderen Höreindruck. Überall dort, wo das Interesse von Kommunikatoren an einer spezifischen Wirkung von Botschaften sehr hoch ist, wie zum Beispiel in der Werbung, sollte der Einsatz bestimmter Stimmen daher sehr genau überlegt werden.[1]

Werbung will im allgemeinen Aufmerksamkeit erregen und Aktivierung erreichen. Es sollen aber auch Glaubwürdigkeit vermittelt werden, Kompetenz ausgedrückt, sowie Gefühle von Vertrautheit und Sympathie aufgebaut werden. Die forschungsleitende Frage der vorliegenden Arbeit ist, ob bei den in Werbespots verwendeten Stimmen dementsprechende Kompetenz- und Aufmerksamkeitssignale identifiziert werden können.

Basierend auf bereits vorliegenden Untersuchungsergebnissen der Stimmforschung lassen sich grundsätzlich zwei verschiedene Stimmtypen, die zur Erreichung dieser Kommunikationsziele verwendet werden, unterscheiden:

- Um Aufmerksamkeit zu erregen werden im allgemeinen Stimmen verwendet, die Wachsamkeit, eventuell auch Alarmbereitschaft, signalisieren. Diese sind eher hoch und werden dadurch auch als lauter wahrgenommen.
- Um Kompetenz zu vermitteln und Vertrauen herzustellen werden hingegen tiefe Stimmen, eventuell mit einem mäßigen Flüsteranteil, zum Einsatz gebracht.

In der vorliegenden Arbeit soll nun versucht werden, bei Stimmen in deutschsprachigen Fernsehwerbespots die zwei oben genannten Kommunikationsziele mittels der entsprechenden physikalischen Parameter nachzuweisen.

Es werden folgende forschungsleitende Fragen formuliert:

- Können Stimmen in Werbespots identifiziert werden, die anhand bestimmter messbarer physikalischer Parameter und prosodischer Merkmale dem Kommunikationsziel der Erregung von Aufmerksamkeit und Aktivierung entsprechen?
- Können Stimmen in Werbespots identifiziert werden, die messbare Signale der Kompetenz aussenden und die dadurch dem Kommunikationsziel Vertrauen zu erwecken zuordenbar sind?

---

[1] Diese Notwendigkeit ist PR- und Werbefachleuten grundsätzlich auch bewusst. Vgl. dazu Mayr, Nora (2006).

Grundsätzlich ist die Bearbeitung dieser Fragestellung erkenntnisoffen. Derzeit gibt es keine der Autorin vorliegenden relevanten Forschungsergebnisse, die eine spezifische Hypothesenbildung rechtfertigen würden. Da dieser konkrete Forschungsbereich bisher im deutschsprachigen Raum nicht bearbeitet wurde, kann auch auf keine theoretischen Grundlagen oder standardmäßigen Werkzeuge zurückgegriffen werden.

An dieser Stelle muss erklärend darauf hingewiesen werden, dass Stimm- und insbesondere Prosodieforschung sprachspezifisch ist und daher Forschungsergebnisse zu anderen Sprachräumen nur vergleichend von Interesse für die vorliegenden Forschungsfragen sein könnten. Dementsprechend wurde auch weitgehend auf die Verwendung von englischsprachiger Forschungsliteratur verzichtet.

Ein weiterer Grund dafür, dass dieses Themenfeld bisher nur von wenigen und sehr spezialisierten Forschungsgruppen bearbeitet wurde, liegt am hohen technischen und rechnerischen Aufwand, der für die komplexen Signalanalysen gesprochener Sprache notwendig ist. Die in der vorliegenden Arbeit durchgeführten Berechnungen wären zum Beispiel vor zehn Jahren in dieser Qualität noch nicht möglich gewesen.

Für die Wirtschaft ist die Stimm- und Prosodieforschung durchaus von großer Relevanz. Insbesondere die Stimm- und Spracherkennung oder die Generierung von möglichst natürlich wirkenden Computerstimmen sind höchst aktuelle Forschungsgebiete der Elektronikindustrie. Die Ergebnisse der industriell betriebenen Forschung sind jedoch nicht in wissenschaftlicher Literatur öffentlich zugänglich oder nachzulesen, allerdings sind sie zum Beispiel bei Smartphones, Navigationssystemen von Autos oder in Sicherheitssystemen anwendbar.

Die vorliegende Arbeit ist in insgesamt sieben Abschnitte mit teilweise mehreren Unterkapiteln gegliedert. Im den folgenden zweiten Abschnitt werden die relevanten wissenschaftlichen und theoretischen Grundlagen dargestellt. Einleitend wird ein Überblick über die Physiologie der menschlichen Stimme und Stimmgebung sowie des Hörorgans und der auditiven Wahrnehmung gegeben. Danach werden die Zusammenhänge und Wechselwirkungen von Stimme und Sprache, von Stimmeigenschaften und Sprechweise, der Prosodie, aufgezeigt. Der dritte Teil dieses Abschnitts ist dem theoretischen kommunikationswissenschaftlichen Rahmen dieser Arbeit gewidmet. Der kommunikative Prozess wird in Bezug gestellt zum Begriff der Signalübertragung sowie zur mathematischen Theorie der Kommunikation nach Shannon.

Im dritten Abschnitt werden einleitend die strategischen kommunikativen Ziele der Werbung, Erregung von Aufmerksamkeit und Vermittlung von Kompetenz, erläutert. Im Anschluss daran werden auf der Basis bereits vorhandener Forschungsergebnisse jene akustischen Parameter besprochen, die für die Analyse der prosodischen Merkmale von Erregung und Kompetenz von Relevanz sind.

Im vierten Abschnitt wird zu Beginn der Auswahlprozess von zur Stimmanalyse geeigneten Fernsehwerbespots beschrieben. Danach werden die Berechnungen folgender akustischer Parameter der verwendeten Stimmen dargestellt und erläutert: die Sprechgeschwindigkeit und die Grundfrequenz mit ihren statistischen Kennzahlen Mittelwert und Median als Parameter der Lage, sowie Standardabweichung (Varianz) und Interquantilsabstand (10-90%) als Parameter der Streuung.

Im fünften Abschnitt werden die Ergebnisse der oben genannten Berechnungen hinsichtlich der forschungsleitenden Fragen analysiert und besprochen. Im darauf folgenden Abschnitt werden auf Basis der Erkenntnisse der vorhergehenden Arbeitsschritte weiterführende Forschungsfragen formuliert. Im siebenten und letzten Abschnitt wird die gesamte Arbeit zusammengefasst.

2. Physiologische und Theoretische Grundlagen

Zum besseren Verständnis der Messungen der Stimmparameter im nachfolgenden dritten Abschnitt wird einleitend ein kurzer Überblick über das Phänomen der menschlichen Stimme gegeben. Es werden die physiologischen Grundlagen der Stimme und Stimmgebung dargelegt, sowie der Einfluss körperlicher Veränderungen und emotionaler Zustände auf die Stimme erklärt. Danach werden, nach einer kurzen Einführung in die Physiologie des Hörorgans, Aspekte der auditiven Wahrnehmung erläutert. Das nächste Kapitel ist den Zusammenhängen und Wechselwirkungen von Stimme und Sprache, von Stimmeigenschaften und Sprechweise, der Prosodie, gewidmet. Der dritte und letzte Teil dieses Abschnitts befasst sich mit dem theoretischen kommunikationswissenschaftlichen Rahmen dieser Arbeit. Der kommunikative Prozess wird in Bezug gestellt zum Begriff der Signalübertragung sowie zur mathematischen Theorie der Kommunikation nach Shannon (1949).

2.1. Die menschliche Stimme und die Stimmgebung

Die menschliche Stimme als „das intimste und gewiss das ausdrucksstärkste Mittel, das uns in der Kommunikation zur Verfügung steht“ (Eckert/Laver 1994:1) ist untrennbar verbunden mit der individuellen Physiologie des Körpers. Sie kann nach Westphal (2002:44) auch treffend als „Spur des Körpers in der Sprache“ bezeichnet werden.

Physiologisch gesehen sind für die Stimmgebung, also die Erzeugung von Tönen und Lauten, primär der Vokaltrakt bzw. das „Ansatzrohr“ oder auch verdeutlichend „Resonanzrohr“ (das ist der Bereich von der Mundöffnung bis zur Stimmritze), der Kehlkopf mit der Stimmritze und den Stimmlippen sowie die Atmungsorgane zuständig.[2] (Siehe Abbildung 1.)

Der Vokaltrakt ist bei erwachsenen Männern durchschnittlich 17 bis 18 cm lang, bei erwachsenen Frauen 16 bis 17 cm. Seine Form beeinflusst maßgeblich den Klang der Stimme. Die Enge oder Weite des Rachens, Lage und Spannung der Zunge, die Stellung des weichen Gaumens (aufgrund der Beweglichkeit auch Gaumensegel genannt) gehören zu den wichtigsten Faktoren, durch die die Stimme ihren individuellen Charakter erhält. Auch die Erzeugung der sprachlichen Laute erfolgt in diesem Bereich: Konsonanten entstehen insbesondere durch die Bewegung der Lippen und verschiedene Stellungen der Zunge, Vokale durch Umformungen des gesamten Ansatzrohres. Sogenannte „nasale“ Laute entstehen durch das Herunterhängen des Gaumensegels (Velum), wie auch in Abbildung 1 dargestellt. Wird dieses jedoch gehoben und liegt dadurch an der hinteren

[2] Trojan (1975:36ff) weist darauf hin, dass diese drei Bereiche nicht isoliert verstanden werden dürfen sondern eng zusammenarbeiten und sich, meist durch antagonistische Muskelfunktionen, gegenseitig beeinflussen. Um dies zu verdeutlichen unterscheidet er sechs Funktionskreise: Atmung, innere sowie äußere Kehlkopfmuskulatur, Rachen- und Schlundkopfmuskulatur, Ventile und Bewegung des Unterkiefers.

Rachenwand an, so ist der Nasenraum vom Rachenbereich getrennt und können sich dadurch weder Luft noch Schall in diesem Bereich bewegen. Dieser Unterschied ist deutlich hörbar.

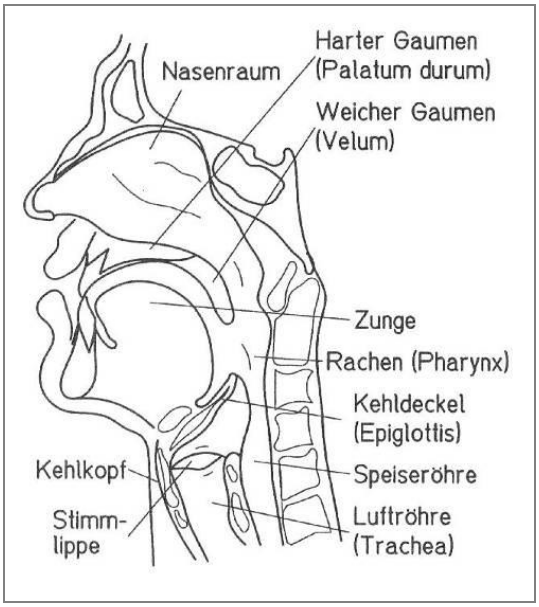


Abbildung 1: Schematischer Querschnitt durch das menschliche Sprechorgan[3]

Der Kehlkopf (Larynx) besteht aus Muskeln und Knorpeln, die mit Gelenkbändern verbunden sind. Diese sind großteils mit Schleimhaut überzogen. Unterhalb des Kehldeckels (Epiglottis) ist die Stimmritze (Glottis) mit den zwei Stimmlippen (Labium vocale). Die freien Ränder der Glottis werden auch Stimmbänder (Plica vocalis) genannt. Die Kehlkopfmuskeln bewegen nicht nur den Kehlkopf, sie öffnen und schließen auch die Glottis und spannen die Stimmbänder.

Bau und Beweglichkeit, sowie Länge und Breite der Stimmlippen sind abhängig vom hormonellen Status. Bei Frauen beträgt die Länge der Stimmlippen 1,3 bis 2 cm, bei Männern 1,7 bis 2,4 cm. Wenn bei der Phonation, also der Erzeugung von Lauten, die Stimmlippen schwingen, misst man die Anzahl dieser Schwingungen pro Sekunde mit der Einheit Hertz (Hz)[4]. Je kürzer die Stimmlippen sind, desto schneller schwingen sie und desto höher sind die erzeugten Laute. Bei Frauenstimmen dauert ein Phonationszyklus beim Sprechen 6,7 ms bis 4,6 ms. Dies entspricht 150 bis 220 Hz bzw. Schwingungen pro Sekunde.[5] Bei Männern schwingen die Stimmlippen zwischen 10 ms und 5,6 ms, was einer Tonhöhe zwischen 100 bis 180 Hz[6] entspricht.[7]

---

[3] Abbildung 1 entnommen aus: Terhardt (1998: 32).

[4] Diese Frequenzeinheit ist nach dem deutschen Physiker Heinrich Rudolf Hertz (1857-1894) benannt.

[5] Wenn in der Musik der Kammerton „a^1“ auf 440 Hz gestimmt ist, entspricht die Tonhöhe von 220 Hz einem „a“, das eine Oktave tiefer liegt. 150 Hz liegen wiederum etwas weniger als das Intervall einer Quinte tiefer als „a“ zwischen den Tönen „d“ und „dis“. (Zum Intervall der Oktave siehe auch Fußnote 29 in Kapitel 2.2.)

[6] Wenn „a^1“ auf 440 Hz gestimmt ist, ist die Tonhöhe von 100 Hz in der Musik etwas tiefer als ein „A“, 180 Hz ist etwa das Intervall einer Sext höher und etwas höher als ein „f“. Zwischen dem höheren Wert der Frauenstimme und dem tieferen Wert der Männerstimme liegt somit etwa eine Oktave.

Die Angaben über die Schwingungsfrequenzen bei den Sprechstimmen von Frauen und Männern sind nicht einheitlich. Beispielsweise findet man bei Terhardt (1998:33) als Mittelwert der Sprechstimme (auch Stimmgrundfrequenz oder kurz f_0 genannt) für Männer 120 Hz, für Frauen 240 Hz, was ebenfalls eine Oktave Differenz bedeutet, jedoch einen tieferen Wert für Männer und einen höheren für Frauen bedeutet. Goldstein (2002:485) gibt für Männer ebenfalls 120 Hz an, für Frauen jedoch nur 210 Hz.[8] Für diese Abweichungen der Werte gibt es mehrere Gründe: erstens die Komplexität der Messungen, auf die im nächsten Abschnitt dieser Arbeit noch ausführlicher eingegangen werden wird, zweitens die sprach- sowie kulturspezifischen Unterschiede bei Sprechstimmen[9]. Ohne auf diese Thematik hier ausführlicher eingehen zu können, soll doch darauf hingewiesen werden, dass die angegebenen Werte nicht als absolut sondern vielmehr als durchschnittliche Richtwerte zu verstehen sind.

Die Mitarbeit am Sprechvorgang ist eigentlich eine sekundäre Funktion des Kehlkopfs. Primär ist er für den Verschluss der Luftröhre (Trachea), die vom Kehlkopf zu den Bronchien führt, zuständig. Dadurch verhindert er das Eindringen von Flüssigkeiten und Nahrungsteilchen oder hilft bei der Akkumulation von Luft in der Lunge, was zum Beispiel für das Husten wichtig ist. Die Weite der Glottis ist jedoch abhängig von der Atmung. Je ruhiger diese ist, desto schmäler ist die Glottis. (Siehe Abbildung 2.)

Für die Phonation ist der Luftstrom unabdingbar. Erstens benötigen Schallwellen die Teilchen der Luft für ihre Verbreitung,[10] zweitens hängt die Lautstärke von der Stärke des Luftstroms ab, drittens dient der Luftstrom als Rückstellkraft für die Bewegung der Stimmlippen.

Bedenkt man, dass die Stimmlippen einer Frau beim Sprechen etwa 220 mal pro Sekunde schwingen können, bei hohen Tönen einer Sängerin sogar mehr als tausend mal[11], dann wird verständlich, dass es verschiedener Hilfsmittel bedarf um diese Geschwindigkeiten zu bewältigen. „Würden die Stimmlippen nur durch Muskelkraft bewegt werden, wie z. B. beim Hin- und Herbewegen der Hand, so würden wir damit keine hörbaren Töne hervorbringen können.“

---

7 Diese Werte sind entnommen aus: Mayer (2010) o. S.

8 Diese Angaben sind bei Goldstein (2002: 485) eigentlich vertauscht. Er gibt 120 Hz für Frauen und 210 Hz für Männer an. Dies ist jedoch mit Sicherheit unrichtig und kann auf einen Schreib- oder Übersetzungsfehler zurückgeführt werden.

9 Eckert/ Laver (1994) zitieren eine australische Studie, bei der Aufnahmen von Frauenstimmen seit den 1940er Jahren verglichen wurden. Dabei wurde festgestellt, dass die Stimmen in den letzten Jahren immer tiefer wurden. Begründet wird dies damit, dass Frauen heute größer gewachsen sind und daher auch längere Stimmbänder haben sowie mit sozialen Faktoren: „Es ist heute gesellschaftlich eher akzeptabel, nicht so weiblich zu klingen. Eine tiefere Stimme wird mit Reife und Autorität in Verbindung gebracht.“ (1994:36)
Westphal (2002:172) kommt bei einem Vergleich von Frauenstimmen zwischen 1908 und 1997 ebenfalls zu der Erkenntnis, dass die Stimmen der Frauen vor den 1970er Jahren wesentlich höher waren als in der Zeit danach. Sie weist allerdings darauf hin, dass hier auch ein Zusammenhang mit geänderten technischen Bedingungen (Aufnahmetechnik, Mikrofone) bestehen kann.
Ebenfalls dazu siehe auch Geissner (2004).

10 Dieser Vorgang ist gut erklärt bei Eckert/ Laver (1994:179f).

11 Das entspricht etwa der Höhe eines „c^3“.

(Eckert/Laver 1994:52) Insgesamt wirken daher drei verschiedene Rückstellkräfte:[12] die Muskulatur des Kehlkopfs; das elastische Gewebe der Stimmlippen, das die Tendenz hat, sich immer in die Ausgangsposition der geschlossenen Stimmritze zurückzubewegen; sowie das „aerodynamische Paradoxon“[13]. Auf den Phonationsvorgang übertragen bedeutet dieses, „daß die Luft, die aus der Lunge durch die Luftröhre nach außen gepreßt wird, an dem Glottisspalt, der einen geringeren Querschnitt aufweist als die Luftröhre, auf eine Verengung trifft. Die Fließgeschwindigkeit der Luft nimmt zu, während an der Glottis ein Unterdruck entsteht, der die Stimmlippen zusammenzieht und die Glottis verschließt (Bernoullikräfte). Unterhalb der verschlossenen Glottis baut sich dann wieder ein Druck auf, der die Stimmlippen rasch wieder auseinander sprengt, die angestaute Luft kann durch die Glottis entweichen, was wiederum zu einem Unterdruck und dem Verschluß der Glottis führt.“ (Mayer 2010: o. S.)

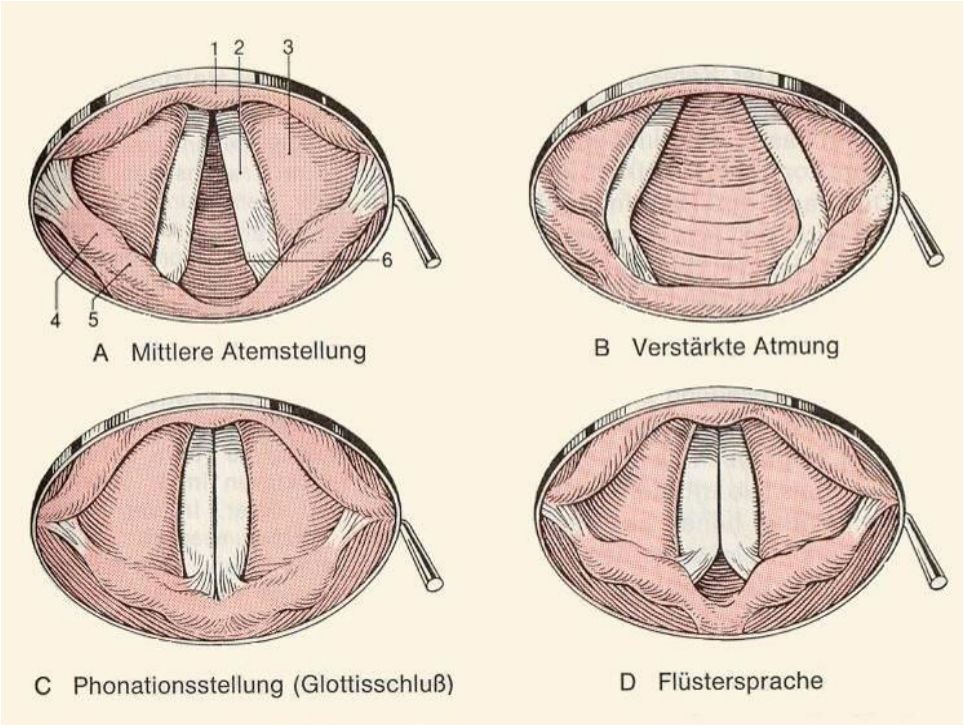


Abbildung 2: Ansicht der Glottis bei einer Kehlkopfspiegelung[14]

(Im Bild oben: 1: Die Begrenzungen der Stimmritze: 2: Stimmfalten; 3: Taschenfalte; 4: Tuberculum coneiforme; 5: Tuberculum corniculatum; 6: Processus vocalis)

Es sind jedoch nicht nur die physikalischen und organischen Faktoren der Luftzufuhr, die für die Stimmgebung wichtig sind, auch die Qualität der Atmung (die Geschwindigkeit, Gleichmäßigkeit

[12] Vgl. dazu ausführlicher Eckert/ Laver (1994:51f).
[13] Dieses besagt, dass, wenn die Geschwindigkeit eines Gases oder einer Flüssigkeit so groß ist, dass der statische Druck unter den Atmosphärendruck sinkt, die Strömung eine ansaugende Wirkung erhält. Dieses Phänomen wird nach seinem Entdecker, dem Mathematiker und Physiker Daniel Bernoulli (1700-1782), Bernoullisches Gesetz von 1738 genannt.
[14] Abbildung 2 entnommen aus Leonhardt (1973:121).

und Intensität) spielen eine wesentliche Rolle für die Stimme. Williams/Stevens (1981, in: Scherer 1982:309) weisen darauf hin, dass Veränderungen des Rhythmus und der Frequenz der Atmung den Atemdruck beeinflussen und dadurch auch die Grundfrequenz der Stimme sowie die akustischen Charakteristika der Konsonanten verändert werden.[15] Die Art und Weise der Atmung (langsam oder schnell, regelmäßig oder stoßweise, tief oder flach) ist einerseits determiniert durch die körperliche Befindlichkeit und die Notwendigkeit der Sauerstoffversorgung des Organismus, andererseits auch durch den emotionalen Zustand: „Auch die leichteste Veränderung beim Atmen wirkt sich unweigerlich auf die Stimme aus. Und die leichteste Veränderung des Gefühlszustandes wirkt sich auf die Atmung aus." (Eckert/Laver 1994:169)

Besonders auffällig sind emotionale Erregungszustände, die zu körperlichen Reaktionen führen, die wiederum unüberhörbar für die Stimme von Bedeutung sind. Es sind dies vor allem beschleunigte oder unruhige Atmung, erhöhte Muskelspannung sowie Veränderungen an den Schleimhäuten. Diese Veränderungen wiederum „resultieren in der Modifikation der Artikulationsrate, sowie der akustischen Charakteristika von Vokalen und der Variationsbreite der Frequenzen bei der Vibration der Stimmlippen." (Williams/Stevens 1981, in: Scherer 1982:308)

„Ekel und Angst, aber auch leichte Gefühle der Unsicherheit führen zu Anspannung der Muskeln und zur Verengung von Resonanzräumen im stimmlichen Bereich. Gefühle der Sicherheit und des Wohlbefindens führen dagegen zu Entspannung und Erweiterung der Resonanzräume. Gefühlsmäßige Erregung führt meist zu erhöhter Stimmlage." (Eckert/Laver 1994:11)

Doch nicht nur Atemfrequenz und Muskeltonus stehen in Verbindung mit emotionalen Zuständen und finden dadurch Ausdruck im Stimmklang, auch Funktionen von Magen und Darm beeinflussen direkt die Weite des Rachens und damit wieder den Vokaltrakt. Ist der Rachen geweitet, entspricht dies der Nahrungsaufnahme, die im allgemeinen mit Lustgefühlen verbunden wird, und der Peristaltik, der absteigenden Zusammenziehung und Erschlaffung von Magen und Darm. Rachenenge hingegen hängt mit Unlustgefühlen und der Antiperistaltik, dem Brechakt, zusammen. Dabei werden die hohen Frequenzbereiche der Stimme weniger und es entstehen auch Zusatzgeräusche, die wie ein „ch" aus dem Wort „Ach" klingen. (Trojan 1975, in: Scherer 1982:59ff)

In den zwei folgenden Sonagrammen (Abbildung 3) ist der akustische Unterschied von Rachenweite und -enge optisch gut erkennbar. Beim ersten Bild ist der Satz, „Pfui Teufel, das schmeckt ja ekelhaft!", mit Rachenenge gesprochen. Der Geräuschcharakter wird deutlich durch die eher fleckigen, undeutlichen Wellenmuster. Besonders auffällig ist der kaum sichtbare letzte Vokal „a" im Wort „ekelhaft". Im zweiten Sonagramm sieht man denselben Satz mit Rachenweite gesprochen. Die markanten und gleichmäßigen Wellenlinien, die den harmonischen Teiltönen der Vokale entsprechen, zeugen von einem offenen Vokaltrakt, in dem ausreichend Platz für die

[15] Siehe dazu auch Eisinger (2002), der in seiner Diplomarbeit die Veränderungen der Stimmgrundfrequenz von Soldaten bei psychischen und physischen Belastungen untersucht hat.

Entstehung und Entwicklung einer klingenden Stimme ist. Hier sind die Vokale ausgeprägt, gut sichtbar und die Sprache dadurch auch verständlich und deutlich.

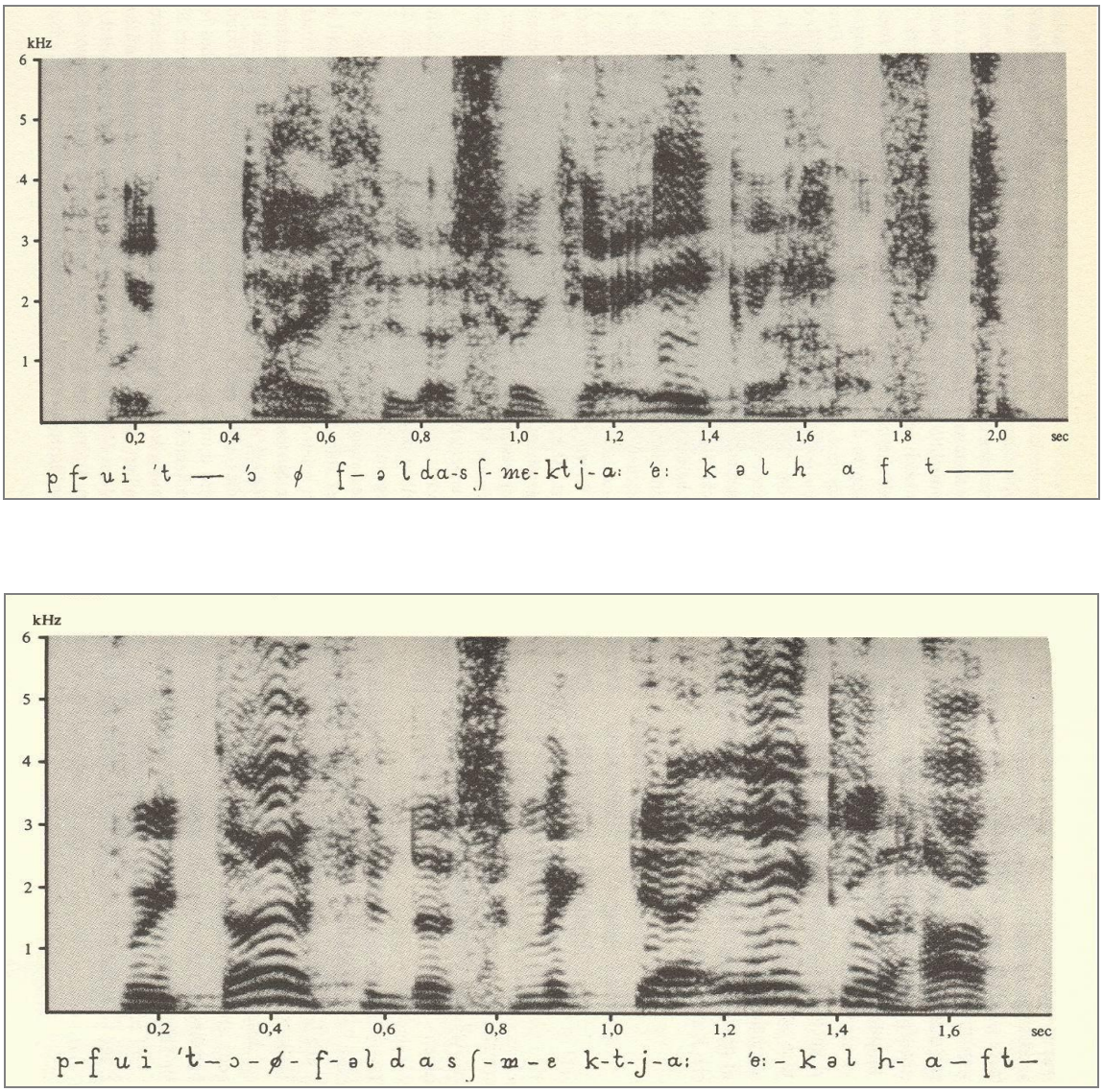


Abbildung 3: Sonagramme des Satzes „Pfui Teufel, das schmeckt ja ekelhaft!“,
im oberen mit Rachenenge, im unteren mit Rachenweite gesprochen.[16]

[16] Abbildung 3 entnommen aus: Trojan (1975) in: Scherer (1982:64f).

2.2. DAS MENSCHLICHE HÖRORGAN UND DIE AUDITIVE WAHRNEHMUNG

Das menschliche Ohr oder Hörsystem ist ein hoch komplexes Sinnesorgan, das mit einigen Superlativen aufwarten kann: dem härtesten Knochen des Schädels (das Felsenbein[17]), den kleinsten Knochen (die drei Gehörknöchelchen, Hammer, Amboss und Steigbügel, im Mittelohr) sowie den kleinsten Skelettmuskeln (die Mittelohrmuskeln bei den Gehörknöchelchen) des Körpers. Desweiteren ist das Ohr das am frühesten entwickelte Sinnesorgan: Bereits rund vier-einhalb Monate nach der Befruchtung der Eizelle ist das eigentliche Hörorgan (das Innenohr mit der Cochlea und dem Vestibulärorgan) des Embryos in seiner endgültigen Größe fertig ausgebildet. Von diesem Zeitpunkt an ist der Fötus fähig Geräusche und Stimmen wahrzunehmen.[18] Funktional von der Cochlea weitgehend unabhängig ist das für den Gleichgewichtssinn zuständige Vestibulärorgan, das zu diesem Zeitpunkt ebenfalls schon aktiv ist.

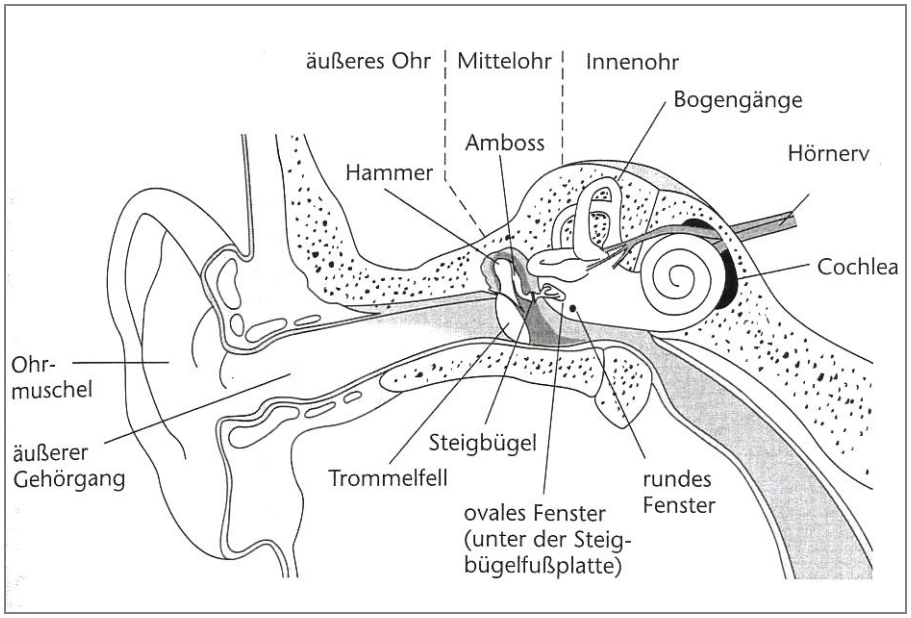


Abbildung 4: Das Ohr mit seinen drei Teilen: äußeres Ohr, Mittelohr, Innenohr[19]

In Abbildung 4 ist ein schematischer Schnitt durch ein rechtsseitiges Hörorgan mit seinen drei Hauptteilen zu sehen. Das äußere Ohr besteht aus den Ohrmuscheln (Auricula, Pinna) und dem äußeren Gehörgang (Meatus acusticus ext.), der bis zum Trommelfell (Membrana tympani) reicht. Mit Hilfe des Ohrenschmalzes ist dieser Bereich vor allem für den Schutz der empfindlichen inneren Teile des Ohres zuständig. Bei der Lautwahrnehmung werden darüber hinaus aufgrund einer Resonanzfunktion Frequenzen zwischen 2.000 und 4.000 Hz leicht verstärkt. (Vgl. Goldstein

[17] Das Felsenbein (Pars petrosa) ist Teil des Schläfenbeins und beherbergt vor allem das Innenohr und den inneren Gehörgang. Ausführlicher dazu siehe zum Beispiel Klingebiel (2002:9f).
[18] Zum pränatalen Hören siehe auch ausführlicher: Berendt (1989), Seimer (2006), Tomatis (1981, 1987, 2003) sowie Westphal (2002:105ff).
[19] Abbildung 4 entnommen aus: Goldstein (2002:389).

2002:388) Das Mittelohr (Auris media) reicht vom Trommelfell bis zum ovalen Fenster der Cochlea. Es ist ein kleiner Hohlraum, auch Paukenhöhle genannt, der die Gehörknöchelchen (oder auch Mittelohrknöchelchen) enthält. Diese werden nach ihrer Form als Hammer (Malleus), Amboss (Incus) und Steigbügel (Stapes) bezeichnet. Mit ihnen verbunden sind die Mittelohrmuskeln, die in der Abbildung jedoch nicht eingezeichnet sind. Das Innenohr (Auris interna) besteht aus der flüssigkeitsgefüllten Cochlea (Schnecke) und dem bereits erwähnten Vestibulärorgan mit den Bogengängen, das für den Gleichgewichtssinn und die Wahrnehmung von Lage und Beschleunigung zuständig ist. Die Cochlea ist eine sehr komplexe schneckenförmige Struktur, in der auch die Haarzellen[20] angesiedelt sind, die wiederum über Synapsen mit den Fasern des Hörnervs verbunden sind.[21]

Treffen nun Schallwellen an das Ohr wird das Trommelfell in Schwingungen versetzt, die wiederum an den Hammer und die nächsten zwei Gehörknöchelchen weitergegeben werden. Über die Platte des Steigbügels werden sie dann auf die Membran des ovalen Fensters übertragen. Dabei wird der Schalldruck vom Trommelfell bis zur wesentlich kleineren Steigbügelpatte (das Größenverhältnis ist ungefähr 17:1) um etwa den 22fachen Wert verstärkt.[22] Ist der Schalldruck allerdings zu groß, kontraktieren die Mittelohrmuskeln und dämpfen dadurch die Schwingungen. Terhardt (1998:56) weist jedoch darauf hin, dass aufgrund des Umweges über mehrere neuronale Schaltstellen „dieser Regelkreis eine Zeitkonstante von der Größenordnung 100 ms“ hat und daher kein Schutz vor kurzen Schalldruckimpulsen mit hoher Amplitude, wie zum Beispiel Knallen, gegeben ist.

Die Schwingungen der Membran des ovalen Fensters übertragen sich im weiteren Verlauf auf die Flüssigkeit (Lymphe) in der Cochlea. Dadurch werden die Haarzellen mit den an ihren Enden befindlichen Sinneshärchen (Stereozilien) in Bewegung versetzt, es werden bioelektrische Signale erzeugt und diese Impulse über die Nervenfasern und den Hörnerv (Nervus cochlearis) an das Gehirn weitergeleitet, wo die neuronalen Signale dann decodiert werden. Die Schallwellen werden hingegen über die Cochlea wieder in das Mittelohr ausgeleitet.[23]

Bereits in der Cochlea werden Schallsignale durch die Haarzellen nach ihrer Frequenz analysiert und zerlegt, das heißt in anderen Worten, dass Tongemische, Geräusche und Klänge in ihren

---

[20] Der gesunde Mensch hat pro Ohr etwa 3.500 innere und 12.000 äussere Haarzellen. Vgl. dazu: Goldstein (2002:391), http://www.uni-duesseldorf.de/MedFak/mai/teaching/content/neuroanatomie/index.php?kap=18 (11.8.2013) sowie http://www.biologie-online.eu/ohr.php (11.8.2013). Terhardt (1998:58) gibt jedoch gesamt ungefähr 25.000 Haarzellen (davon 3.500 bis 4.500 innere Haarzellen) an. Hier scheint es sich jedoch um einen Schreibfehler zu handeln, da keine Literatur gefunden werden konnte, die diese Angabe bestätigt.

[21] Über das Innenohr, die Anatomie und Funktion der Cochlea siehe ausführlicher zum Beispiel: Goldstein (2002:390ff).

[22] Goldstein (2002:390) weist darauf hin, dass bei einer Schädigung der Gehörknöchelchen der Schall auch direkt ins Innenohr geleitet werden kann. Damit die Betroffenen dann jedoch genauso gut hören können, muss der Schalldruck um einen Faktor zwischen 10 und 50 vergrößert werden.

[23] Die Cochlea ist in drei Gänge geteilt. Der Weg der Schallwellen geht über die Scala vestibuli hinein, führt über die Scala media zur Erregung der Sinneszellen und kehrt anschließend über die Scala tympani wieder zum ovalen Fenster und zum Mittelohr zurück. (Vgl. dazu auch Klingebiel 2002.)

Teiltönen[24] erkannt und verarbeitet werden.[25] Im auditorischen Cortex des Gehirns findet dann die Weiterverarbeitung und Interpretation der in der Cochlea analysierten Frequenzen statt.

Die Wahrnehmungskapazität des Hörorgans ist bei gesunden Menschen wesentlich leistungsstärker als die des Sehsinns: „Der Wahrnehmungsspielraum – der *range* – unseres Auges ist etwa eine Oktave breit: von Violett (380 Nanometer) bis Purpur (760 Nanometer) – den beiden Farben am jeweils äußersten Ende der sichtbaren Farbskala – verdoppelt sich gerade die Wellenlänge, also eben eine Oktave. Aber wir können in einem *range* von rund zehn Oktaven hören. (…) Unser Auge braucht 20/1000 Sekunden, um zwei aufeinanderfolgende Reize noch unterscheiden zu können, das Ohr lediglich 3/1000 Sekunden. Das Ohr also ist fast siebenmal schneller." (Berendt 1998:75)

Dieser hoch spezialisierte Prozess des Hörens ist auch ein sehr subjektiver: Erstens werden die Schallsignale durch Form und Größe des Körpers, Kopfes sowie der Ohrmuscheln verändert. Zweitens ist die Qualität der Frequenzanalyse von Anzahl und Zustand der Haarzellen abhängig. Drittens ist auch die Größe des auditorischen Cortex variabel und wird durch Erfahrung und Übung beeinflusst.[26] Darüber hinaus ist auch die Interpretation von wahrgenommenen Schallereignissen subjektiv und nicht unbedingt ident mit physikalisch gemessenen Werten. Beispielsweise hängt die geschätzte Dauer von akustischen Signalen von deren Schalldruckpegel sowie deren Frequenz ab. Je lauter oder je höher der gehörte Tonimpuls ist, als umso länger dauernd wird er empfunden.[27]

Desweiteren korrelieren Frequenz und Schalldruckpegel eines Signals in der auditiven Wahrnehmung: Je tiefer ein Ton ist, desto höher muss der Schalldruck sein, damit er wahrgenommen wird. Je höher ein Signal wird, desto empfindlicher wird das Ohr. Die höchste Empfindlichkeit ist in dem Frequenzbereich, der für die Sprachwahrnehmung notwendig ist.[28] Hier werden die relativ leisesten Signale wahrgenommen. Töne, die darüber liegen, benötigen hingegen wieder einen zunehmend höheren Schalldruck um gehört zu werden.

Dies ist in Abbildung 5, in der die sogenannte Hörfläche grafisch dargestellt wird, deutlich sichtbar. Auf der x-Achse sind die Frequenzen des Hörbereichs zwischen 20 und 20.000 Hz

---

[24] Jeder Klang und auch jedes Geräusch bestehen aus mehreren Teiltönen. Bei einem Klang sind diese im allgemeinen in einem harmonischen Verhältnis zueinander, bei Geräuschen eher nicht. Einen reinen Ton, der nur aus einer Schwingung bzw. Schallwelle besteht, nennt man Sinuston. Dieser kommt in unserem Alltagsleben normalerweise nicht vor. Sinustöne werden jedoch häufig bei Messungen des Gehörs und auditiven Wahrnehmungsanalysen verwendet. (Siehe dazu auch Kapitel 3.2.2.)

[25] Ausführlicher zu den frequenzselektiven Eigenschaften der Cochlea siehe auch Terhardt (1998:239ff).

[26] Bekannt sind diesbezüglich vor allem Forschungsergebnisse, die zeigen, dass bei Personen, die aktiv über längere Zeiträume Musik betreiben, die auditorischen cortikalen Areale deutlich vergrößert sind. Vergleiche dazu zum Beispiel Goldstein (2002:411).

[27] Vergleiche dazu Terhardt (1998:410ff).

[28] Die Angaben über den Frequenzbereich für das Verstehen von Sprache sind ziemlich unterschiedlich. Während bei Westphal (2002:59) und Goldstein (2002: 380) etwa 300 bis 3.000 Hz bzw. 400 bis 3.000 Hz angegeben sind, spricht Hiltensperger (2004:4) vom Schwerpunkt des Informationsgehaltes der menschlichen Sprache im Frequenzbereich von ca. 500 bis ca. 4.000 Hz.

eingezeichnet. Dies entspricht auch etwa den weiter oben von Berendt erwähnten zehn Oktaven.[29] Auf der y-Achse ist der Schalldruckpegel in Dezibel (dB) angegeben, sowie rechts von der Grafik als Vergleichswerte auch die Schallintensität in Watt pro Quadratmeter und der Schalldruck in der physikalischen Größe Pascal (Pa). Die Hörfläche, oder auch Hörfeld genannt, liegt zwischen der Ruhehörschwelle, unterhalb der Frequenzen nicht wahrnehmbar sind, und der Schmerzschwelle, ab der unmittelbare Schäden des Hörorgans auftreten können. Oberhalb der Schwelle für Risiken einer Schädigung können ebenfalls bereits temporäre oder permanente Schäden auftreten. Die gestrichelte Linie rechts unten zeigt das markante Ansteigen der Hörschwelle von Personen, die häufig sehr laute Musik hören. Die erwähnte Empfindlichkeit im Sprachbereich ist deutlich erkennbar am Absinken der Linie der Ruhehörschwelle sowie auch der Schwelle für die Risiken einer Schädigung. Bemerkenswert ist allerdings, dass genau dieser Bereich etwas weniger gefährdet hinsichtlich einer dauerhaften Schädigung zu sein scheint, was am Ansteigen der obersten Kurve des Hörfeldes erkennbar ist.

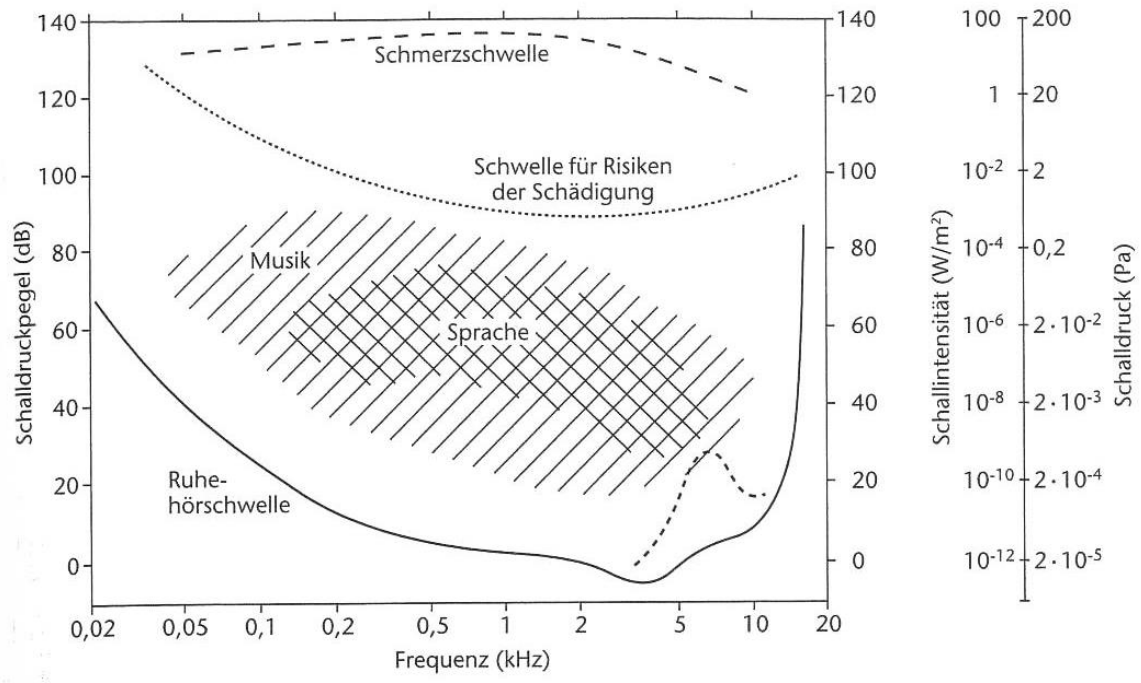


Abbildung 5: Die Hörfläche[30]

Bei Sinneswahrnehmungen wird im allgemeinen zwischen prothetischen und metathetischen Empfindungen unterschieden. Erstere betreffen die Intensität, bei akustischen Signalen also vorwiegend die Lautheit, zweitere die Ortsempfindung, die beim Hörsinn in erster Linie der Tonhöhe entspricht. Terhardt (1998) differenziert die prothetischen Aspekte als Schallstärke und

[29] Der Begriff Oktave bezeichnet eine bestimmte Größe des Intervalls zwischen Tönen bzw. Schallsignalen - und zwar eine Verdoppelung (hinauf) oder Halbierung (hinunter) des Frequenzwertes. Geht man im vorliegenden Fall von der Basis von 20 Hz aus, ist der Wert der zehnten Oktave 20.480 Hz.
[30] Abbildung 5 nach Zwicker/Fastl (1999) in: Goldstein (2002:381).

den damit verbundenen Attributen der Schwankungsstärke und Rauigkeit, sowie Volumen, Schärfe und Klanghaftigkeit.[31] Die metathetische Hörempfindung der Tonhöhe dient in erster Linie der Erkennung von Schallobjekten. Dieser Prozess ist hoch kompliziert - unter anderem deshalb, weil Geräusche, musikalische Töne und auch die menschliche Stimme eine ganze Anzahl von Teiltönen und Tonhöhen enthalten.

> „Die besondere Art auditiver Schallanalyse und -synthese, welche sich in der Tonhöhenwahrnehmung zeigt, ist ein typisches Merkmal der Art und Weise, wie das Gehör seinen natürlichen Zweck erfüllt. Dieser besteht unter anderem darin, die Anzahl der Schallobjekte, welche das aktuelle Ohrsignal verursachen, sowie deren Art herauszufinden. Wegen der Unbestimmtheit der Ohrsignale ist diese Aufgabe niemals eindeutig lösbar. Dies ist - eben weil das Gehör an die real vorgegebenen Bedingungen der akustischen Informationsgewinnung angepaßt ist - der Grund dafür, daß die Tonhöhe eines komplexen Schallsignals niemals völlig eindeutig sein kann.“ (Terhardt 1998:315)

Die Kombination dieser prothetischen und metathetischen auditiven Wahrnehmungen und die Interpretation derselben ergeben zusammen die faszinierenden Fähigkeiten des Hörorgans. Denn dieses ist auf Basis der auralen Frequenzanalyse imstande zwischen primären Schallquellen und Umgebungsgeräuschen zu differenzieren, die Entfernungen von Schallquellen zu schätzen und sie örtlich zu lokalisieren, sowie diese Objekten zuzuordnen. Und mittels der Stimmen von Menschen können nicht nur deren Geschlecht, Alter und Emotionen erkannt sondern darüber hinaus auch deren Sprache verstanden werden.

Die Hauptfunktionen des Hörens, wie in Abbildung 6 dargestellt, können wie folgt zusammengefasst werden:

- Wahrnehmen der Umwelt: Erkennen von Objekten und von größeren Umgebungsbereichen
- Wahrnehmen von Musik
- Wahrnehmen von Sprache

Jeder dieser drei Bereiche, die im Alltag oft auch noch gleichzeitig oder in rasch wechselnder Abfolge auftreten, erfordert vom Hörorgan unterschiedliche Leistungen. Bei der Wahrnehmung der Umwelt geht es primär um das Erkennen und Zuordnen von Hörobjekten und Umgebungs-geräuschen. Hierbei stützt sich das Hirn auch auf bereits abgespeicherte Informationen zu bestimmten akustischen Merkmalen. Bei Musik ist im allgemeinen nicht Horchen sondern Zuhören erforderlich. „Es müssen die Motive, die Rhythmik, die Melodien, die Sequenzen, die Harmonie, die Konsonanzen und Dissonanzen wahrgenommen werden.“ (Goldstein 2002:444) Wobei Goldstein hier von aktivem Zuhören spricht und nicht von Musik, die als Hintergrundbeschallung eingesetzt wird. In diesem Fall wäre sie vielmehr den akustischen Umgebungsbereichen

[31] Siehe dazu ausführlich Terhardt (1998:271ff).

zuzuordnen. Beim Wahrnehmen von Sprache geht es einerseits um das Analysieren und Verstehen von akustischen Sprachsignalen, andererseits um das Erkennen der sprechenden Person.

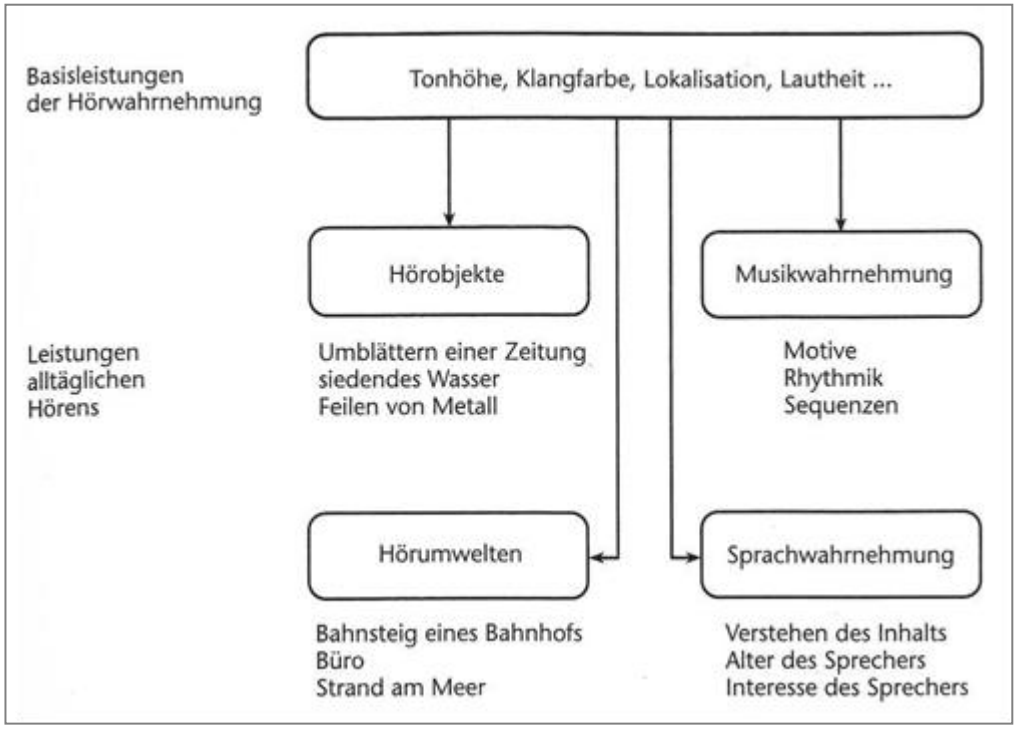


Abbildung 6: Funktionsbereiche alltäglichen Hörens[32]

Abgesehen von diesen genannten Aufgaben ist das Ohr auch der wichtigste Energielieferant für das Gehirn:

> „Das Hirn braucht zum Leben Zucker und Sauerstoff, kann damit allein aber noch lange nicht denken. Für diese Funktion benötigt es eine andere Art von Nahrung: Stimuli, die aus allen Sinnesorganen als Fortleitung elektrischer Potentiale zu ihm gelangen. […] Das hierfür weitaus wichtigste Sinnesorgan ist das Ohr, das ungefähr mit 90 Prozent an der Energiezufuhr zur Hirnrinde beteiligt ist; und dies fast ausschließlich durch den Empfang hoher Frequenzen. In der Schnecke (Cochlea), dem Hörorgan des Innenohrs, befinden sich im Bereich der Wahrnehmung hoher Frequenzen viel mehr Sinneszellen als im Bereich der tiefen. Hohe Frequenzen setzen sich somit in eine unverhältnismäßig größere Zahl von Impulsen um, die eine wahre ‚Aufladung', eine Belebung der kortikalen Tätigkeit bewirken (im EEG sichtbar). Das bedeutet Bewußtsein, Denkfähigkeit, Gedächtnis, Wille usw. - kurz: geistige Wachheit, aber auch Vitalität und Kreativität." (Manassi 2003:17f)

[32] Abbildung 6 entnommen aus: Goldstein (2002:443).

Obwohl diese Funktion des Ohres primär nichts mit der bewussten Rezeption von Botschaften und Signalen zu tun hat, ist sie doch von vitaler Bedeutung. Bemerkenswert ist dabei die Orientierung an hohen Frequenzen. Dies kann hypothetisch auch in Zusammenhang mit der kommunikativen Funktion hoher Stimmen zur Erregung von Aufmerksamkeit, Wachsamkeit und Aktivierung gesehen werden.

2.3. SPRACHE, STIMME UND PROSODIE

Der Begriff der Sprache hat in verschiedenen Kontexten höchst unterschiedliche Bedeutungen, die weit über das allgemeine Verständnis von menschlicher Sprache als kulturspezifischem Kommunikationssystem hinausgehen.[33]

Der Unterscheidung von Saussure (1931/1967) folgend kann die menschliche Sprache grundsätzlich entweder als linguistisches Zeichensystem, „langue", oder als Rede, „parole", betrachtet werden.[34] Die „langue" ist Forschungsgegenstand der Linguistik. Diese beschäftigt sich mit „langue" als vorwiegend statischem System von Subsystemen, wie zum Beispiel mit Morphologie, Syntax oder der Semantik der Sprache. Mit „parole" meint Saussure den individuellen Gebrauch von Sprache in Texten sowie in der mündlichen Rede (Nöth 2000:76). Diese ist Forschungsgegenstand von Sprechwissenschaft und Sprechakttheorie.[35]

Aus kommunikationswissenschaftlicher Sicht wird Sprache in erster Linie als „Medium symbolisch vermittelter Interaktion", als „Instrument zur zwischenmenschlichen Verständigung" (Burkart 2002:77f) gesehen. Denn „alle Kommunikation bedarf eines Mittels oder Mediums, durch das hindurch eine Nachricht […] aufgenommen wird". (Graumann 1972:1182, in: Burkart 2002:36) Ein Medium kann dabei personale Vermittlungsinstanz sein oder auch ein technisches Hilfsmittel. Pross (1972, in: Burkart 2002:36f) unterscheidet drei verschiedene Formen:

- Primäre Medien: Dies sind die Medien des „menschlichen Elementarkontaktes" (Pross 1972:10, in: Burkart 2002:36). Gemeint sind damit alle leiblichen Ausdrucksmöglichkeiten, verbale wie nonverbale.
- Sekundäre Medien: Dies sind Hilfsmittel für Kommunikatoren wie Zeichen oder Signale sowie alle geschriebenen oder gedruckten Medien (Bücher, Zeitungen, Plakate).

---

[33] Vgl. dazu: Hauser/Chomsky/Fitch (2002).
Zur weiteren und engeren Begrifflichkeit von Sprache siehe auch: Lenke/Lutz/Sprenger (1995).

[34] In der deutschen Sprache gibt es keine klare begriffliche Trennung zwischen geschriebener und gesprochener Sprache. In anderen Sprachen wird hingegen deutlich unterschieden zwischen „langue" und „parole" (franz.), „language" und „speech" (engl.) oder „lingua" und „parola" (ital.).

[35] Zur Sprechwissenschaft, Sprechakttheorie und Handlungstheorie der Sprache vergleiche zum Beispiel: Geissner (1981), Allhoff (1983) und Werlen (1984).

- Tertiäre Medien: Damit sind Kommunikationsmittel gemeint, die bei Kommunikatoren und Rezipienten technische Hilfsmittel erfordern, wie zum Beispiel Telefon, Computer[36] oder Fernsehen.

Obwohl die Stimme bei Pross und Burkart nicht explizit erwähnt wird, so gehört sie doch in jedem Fall zu den leiblichen Ausdrucksmöglichkeiten und kann somit als primäres Medium menschlicher Kommunikation bezeichnet werden.

Sprache kann im Sinne von Pross als primäres und sekundäres Medium angesehen werden, die auch über tertiäre Medien vermittelt werden kann. Stimme hingegen ist nur primäres Medium, das ebenfalls über tertiäre Medien kommuniziert werden kann.

Wird Stimme über ein tertiäres Medium reproduziert, verliert sie jedoch „stark an Gehalt, Inhalt und Information“ (Westphal 2002:197). Dies liegt einerseits an den technischen Qualitäten der Signalübertragung andererseits auch daran, dass aufgrund der synästhetischen Eigenschaften unserer Sinneswahrnehmung gleichzeitig mit dem Hören eine Vielzahl weiterer Informationen verarbeitet wird. Eine natürliche Stimme wird daher anders wahrgenommen als eine, die über ein tertiäres Medium übertragen wird. „Die Stimme im Medium ist örtlich und zeitlich dem Sprechenden entrückt. Was wir bei ihr synästhetisch wahrnehmen, ist den Kontext unserer Situation, nicht den Raum-Zeit-Kontext des Sprechenden selbst.“ (Westphal 2002:197) Es muss daher betont werden, dass im Prozess der Rezeption eine Stimme als primäres Medium nicht automatisch gleichzusetzen ist mit einer Stimme, die über ein tertiäres Medium kommuniziert wird.

Stimme und gesprochene Sprache sind beide primäre Medien, die wiederum einander bedingen und gegenseitig beeinflussen.[37] Denn abgesehen von den individuellen physiologischen Merkmalen der Stimme ist die Art und Weise des Sprechens, also des Gebrauchs der Stimme, einerseits abhängig von situativen andererseits auch von kulturellen und sozialen Faktoren. Somit werden über die Stimme und Sprechweise auch Informationen vermittelt, die über die rein inhaltliche, theoretisch auch schriftliche, Mitteilung einer Botschaft hinausgehen:

> „Die Möglichkeit, dem Gesagten durch die Sprechmelodie und den Klang der Stimme einen bestimmten Ausdruck zu verleihen, hebt die gesprochene Sprache von der Schriftsprache hervor. Die gesprochene Sprache erfüllt mehr kommunikative Funktionen als die nur rein linguistischen. Diese Möglichkeit ist gleichzeitig Notwendigkeit, denn ein Sprecher kann gar nicht verhindern, dass mit jeder Äußerung auch Informationen über ihn selbst preisgegeben werden. Besonders Informationen über das Geschlecht, das Alter und Anzeichen für den gesundheitlichen Zustand (z.B. Heiserkeit, Erkältung)

[36] Aus heutiger Sicht wäre natürlich auch das Internet als tertiäres Medium miteinzubeziehen.

[37] Die Stimme kann grundsätzlich auch ohne des Einsatzes eines sprachlichen Zeichensystems verwendet werden. Doch ist jeglicher Stimmgebrauch, ob Sprechen, Singen, Lallen, Summen, eine Form von individuellem Ausdruck, von monologischer oder dialogischer Kommunikation, und kann daher auch als erweiterte Form einer sprachlichen Mitteilung verstanden werden. Auf diese Formen der vokalen Kommunikation einzugehen, würde hier jedoch zu weit führen. Siehe dazu beispielsweise: Scherer (1982).

> werden dem Hörer als Charakterisierung der Quelle stets mitgeliefert. Aber der Sprecher hat auch vielfältige Möglichkeiten, seinen Sprechausdruck bewusst oder unbewusst zu gestalten und damit seine persönlichen Einstellungen, Haltungen und Emotionen deutlich werden zu lassen.“ (Sendlmeier/Bartels 2005:1)

Diese zusätzlichen Informationen werden zusammenfassend auch als prosodische beziehungsweise paralinguistische Merkmale einer sprachlichen Mitteilung bezeichnet.[38] Nach Enterlein / Bartels / Sendlmeier (2005) werden zur Prosodie alle suprasegmentalen Merkmale gezählt, also all jene, die über ein Segment, einen Laut, hinausgehen.[39] Die Eigenschaften einzelner Laute werden hingegen als segmentale Merkmale bezeichnet. Prosodische Merkmale sind zum Beispiel Tonhöhe, Sprechgeschwindigkeit, Lautheit, aber auch Betonungen, Sprechrhythmus und -melodie sowie die Stimmqualität. Wobei Enterlein / Bartels / Sendlmeier zwischen linguistischen und nicht-linguistischen Funktionen der Prosodie unterscheiden. Zu ersteren zählt zum Beispiel der Satzmodus (interrogativ, deklarativ, imperativ usf.), zweitere sind „vor allem paralinguistische Funktionen wie beispielsweise der Ausdruck von Einstellungen und Absichten, aber auch die unwillkürliche Preisgabe von Informationen über Geschlecht und Alter des Sprechers sowie seinen momentanen emotionalen Zustand“ (Enterlein / Bartels / Sendlmeier 2005:10). Eckert / Laver (1994:26) sowie Standke (1993:9) hingegen definieren paralinguistische Merkmale in erster Linie nur als zeitlich begrenzte, nicht-linguistische Funktionen des Ausdrucks von Emotionen und Stimmungen. Paeschke (2003) wiederum arbeitet mit der begrifflichen Unterscheidung von Makro- und Mikroprosodie. Makroprosodie bezeichnet dabei die bewusste stimmliche Steuerung der Sprache um Satzmelodien oder Tonhöhenakzente zu erzeugen. Dies entspricht der weiter oben beschriebenen linguistischen Funktion der Prosodie. Mikroprosodisch sind hingegen die Merkmale, die nicht bewusst steuerbar sind. Bei beiden sind allerdings emotionsspezifische Ausprägungen zu erwarten: „Einerseits hat der Sprecher zum Ausdruck von Emotionen in der Sprechweise bewusst steuerbare Mechanismen zur Verfügung, andererseits wirken nicht bewusst steuerbare physiologische Veränderungen auf die Stimme und Sprechweise“. (Paeschke 2003:30)

Ob nun die Begrifflichkeiten der Merkmale von stimmlichen Äußerungen nach linguistischen, emotionalen oder bewusst steuerbaren Aspekten kategorisiert werden, in jedem Fall wird anhand dieser unterschiedlichen Definitionen auch die Komplexität dieses Forschungsthemas deutlich. Denn durch die untrennbare Verbindung der sprachlichen Aussage mit der individuellen akustischen Färbung durch die Stimme, den Körper und die Emotionen der sprechenden Person sind unzählbar viele Variationen des Ausdrucks und der Interpretation derselben möglich. Dieser

---

[38] Bis Anfang der 80er Jahre gab es dazu keine eindeutige Terminologie (Scherer 1982:88f). Doch auch in späteren Publikation wird immer wieder auf die uneinheitliche Verwendung der Begriffe hingewiesen (Paeschke 2003:24, Enterlein / Bartels / Sendlmeier 2005:10, Geissner 2004:181, 183).

[39] Eine ausführliche Darstellung der verschiedenen Laute (Vokale, Nasale, Plosive und Frikative) und der spezifischen, teilweise sehr komplexen, Strukturen ihrer Schallsignale findet sich bei Terhardt (1998:183-200).

Vielschichtigkeit wird die pragmatische und alle suprasegmentalen Merkmale umfassende Definition des Begriffs Prosodie (Enterlein / Bartels / Sendlmeier 2005) am ehesten gerecht.

Es sind jedoch nicht nur physiologische und situative Faktoren, die eine sprachliche Mitteilung kennzeichnen, auch soziokulturelle Gegebenheiten spielen bei der Art der Verwendung der Stimme eine Rolle. Sprachen und Dialekte unterscheiden sich nicht nur nach linguistischen Kriterien sondern auch nach prosodischen wie zum Beispiel Nasalität, Sprachmelodie, Stärke der Obertöne, desweiteren ist die Art des Sprechens in bestimmten Situationen kulturabhängig. Beispielsweise sprechen japanische Soldaten, die Autorität zeigen wollen, nicht nur mit lauter sondern auch mit sehr tiefer, rauer Stimme. Genau diese Sprechweise wird jedoch in den USA negativ als böse interpretiert.[40] Amerikanische Männer sprechen hingegen wieder in einer wesentlich tieferen Stimmlage als in Deutschland üblich ist: „Ein amerikanischer Mann, der an der unteren Frequenzgrenze des in Amerika üblichen Bereiches liegt, würde in Deutschland als unnormal tief sprechend eingestuft, was natürlich seine Persönlichkeitsbeurteilung durch Deutsche beeinflussen könnte." (Eckert/Laver 1994:156)

Soziokulturelle Unterschiede können jedoch auch innerhalb einer Sprache beobachtet werden. Eckert verweist dabei auf die Stellung des Kehlkopfs als Mittel der regionalen und sozialen Abgrenzung: „ … im Norden Deutschlands, an der Grenze zu Dänemark, da gibt es ganz weit verbreitet den hochgezogenen Kehlkopf als Merkmal dieser Region." Diese Sprechweise zeigt in Schleswig-Holstein die soziale Zugehörigkeit an: „Je höher der Kehlkopf ist und je enger der Pharynx, desto weiter unten ist die soziale Schicht." (Eckert in: Geissner 2004:193) Anfang des 20. Jahrhunderts galt allerdings diese Art des Sprechens wiederum als stimmliches Merkmal sozial hochstehender Personen: „Wer hoch ist, spricht hoch." (Geissner 2004:194)

Zusammenfassend kann festgehalten werden, dass eine sprechende Person gemeinsam mit der inhaltlichen Botschaft eine Vielzahl an weiteren Informationen über sich selbst, ihren Status und ihre regionale Herkunft kommuniziert. Diese prosodischen Codes können, müssen jedoch nicht, von den Hörenden verstanden werden. Dies sollte bei der Bewertung von Stimmen und gesprochenen Botschaften berücksichtigt werden.

2.4. Kommunikation als Prozess der Signalübertragung: die mathematische Theorie der Kommunikation

Nach Burkart (2002:63f) impliziert jeder Kommunikationsprozess

- einen Kommunikator, das heißt, einen kommunikativ Handelnden, jemanden, der etwas mitteilen will
- eine Aussage, das heißt, die mitzuteilenden Bedeutungsinhalte

[40] Ausführlicher dazu siehe Chaika (1989), zit. in: Eckert/Laver (1994:156).

- ein Medium, das heißt, eine Instanz, die die Aussage transportiert
- einen Rezipienten, das heißt, jemand, der die medial vermittelte Botschaft aufnimmt und deren Bedeutung zu erkennen sucht.

Kommunikation als solche findet nach Burkart nur statt, wenn „Verständigung über die mitgeteilte Aussage zustande kommt“, wenn also Kommunikator und Rezipient die Bedeutung der medial vermittelten Aussage auch „tatsächlich miteinander teilen“ (Burkart 2002:66).

Bei dieser Definition des Kommunikationsprozesses liegt der Fokus auf dem willentlichen, bewussten Akt der Vermittlung von Botschaften und dem Verständnis der Bedeutung derselben. Faktoren, die die Inhalte und deren Interpretation beeinflussen können - wenn man so will, der Metabereich des kommunikativen Prozesses - werden dabei nicht explizit berücksichtigt. Bezieht man jedoch die physiologischen Faktoren des Sprechens und des Hörens sowie die prosodischen Eigenschaften der gesprochenen Sprache in die Analyse des kommunikativen Prozesses mit ein, bedarf es eines detaillierteren theoretischen Modells.

Der Mathematiker Claude E. Shannon entwickelte in den späten 1940er Jahren die Theorie eines Kommunikationsmodells, die auch als Theorie des Prozesses der Signalübertragung verstanden werden kann (Shannon/Weaver 1949, 1998). Diese enthält fünf grundsätzliche Elemente:

- Eine Informationsquelle (information source), die aus einer Anzahl möglicher Botschaften eine erwünschte Botschaft (message) auswählt. Botschaften können geschriebene oder gesprochene Worte oder Zeichen sein, ebenso Musik, mathematisch als Funktionen f(t) in der Zeit t.
- Einen Sender (transmitter), der diese Botschaft in
- ein Signal (signal) verwandelt, das über
- einen Kommunikationskanal (channel) zum
- Empfänger (receiver) gelangt, der diese Signale rückverwandelt und zu ihrem Bestimmungsort (destination) weiterleitet.

Im Kommunikationskanal können Störquellen (noise sources) wirksam sein. Damit sind Noise-faktoren gemeint, durch die die Signale verzerrt oder gestört werden. Diese sind vom Sender nicht beabsichtigt und können auch nicht verhindert werden. (Vgl. Shannon/Weaver 1949, 1998:7f, 33f.)

Forschungsarbeiten, die auf dieser Theorie basieren, beschäftigen sich vor allem mit technischen Aspekten der Nachrichtenübertragung. Badura (2004:17) zählt in dem Zusammenhang beispielhaft folgende Fragen auf:

- Wie lässt sich die Informationsmenge messen?
- Wie lässt sich die Kapazität eines Kommunikationskanals messen?
- Welches sind die Merkmale eines effizienten Kodierungsprozesses bei der Umwandlung von Botschaften in Signale?
- Welches sind die Merkmale der Noisefaktoren?

Dieses naturwissenschaftlich theoretische Modell der Signalübertragung kann auch gut auf den Prozess der sprachlichen Kommunikation übertragen werden. Nach Weaver können zum Beispiel bei einem Gespräch zwischen zwei Menschen das Gehirn des Kommunikators als Informationsquelle, der Sender als der physiologische Stimmapparat des Sprechers bezeichnet werden, die Hörorgane des Rezipienten als Empfänger, dessen Gehirn als der Bestimmungsort: „In oral speech, the information source is the brain, the transmitter is the voice mechanism producing the varying sound pressure (the signal) which is transmitted through the air (the channel).“ (Shannon/Weaver 1949, 1998:7) Dementsprechend sind also die Stimme bzw. die gesprochene Sprache das Signal, das über den Kommunikationskanal der Luft (als Trägerin der Schallwellen) übertragen wird.

Grundsätzliche Fragen der Stimm- und Prosodieforschung, aber auch der rezipientenorientierten Wahrnehmungspsychologie oder Psychoakustik können auf Basis des Kommunikationsmodells von Shannon wie folgt formuliert werden:

(1) Wie kodiert der physiologische Stimmapparat des Sprechenden (transmitter) die ausgewählte Botschaft?

(2) Wie dekodiert das Hörorgan des Rezipienten (receiver) die gesendete Botschaft?

(3) Welche sind die physikalischen Eigenschaften des Signals (der Stimme)?

(4) Welche Noisefaktoren beeinflussen das Signal (die Stimme)?

(5) Wie beeinflussen physiologische Eigenschaften bzw. die individuellen Charakteristika des Sprechapparates (transmitter) die Botschaft?

(6) Wie beeinflussen Merkmale des Signals (der Stimme) die Botschaft?

Frage (1) wird vor allem im Bereich der Medizin (Laryngologie) behandelt. Frage (2) ist Thema der Medizin (Phoniatrie) sowie der Psychoakustik, die sich unter anderem auch mit der neuronalen Verarbeitung von auditiv wahrgenommenen Signalen beschäftigt. Frage (3) behandelt Grundlagen jeglicher Stimmforschung, die sich mit den Eigenschaften der Stimme beschäftigt. Frage (4) betrifft Forschungsgebiete, die sich mit der Übertragung und Aufzeichnung von Signalen beschäftigen, von der Akustik, Tontechnik, bis zur Nachrichten- und Signalübertragung. Im Zusammenhang mit der Kommunikationsforschung sind Fragen, die sich zum Beispiel mit der Bedeutung von Hintergrundmusik in kommunikativen Prozessen beschäftigen, diesem Themenkomplex zuordenbar. Die Fragen (5) und (6) sind Themen der Stimm- und Prosodieforschung wie auch der Kommunikationswissenschaft.

Die Forschungsfragen der vorliegenden Arbeit sind in erster Linie den Fragen (3) und (6) zuordenbar.

2.5. ZUSAMMENFASSUNG

In diesem Abschnitt wurden die physiologischen Grundlagen des Stimmapparates erklärt mit dem Ziel, den nicht verhinderbaren Einfluss des Körpers sowie emotionaler Zustände auf die Stimme zu verdeutlichen. Aus diesen Zusammenhängen ergibt sich ein individueller, unverwechselbarer und einzigartiger Stimmcharakter. Verbunden mit der Sprechweise erhält die gesprochene Sprache somit eine Vielfalt an Ausdrucks- und Interpretationsmöglichkeiten, die weit über diejenigen des geschriebenen Wortes hinausgehen. Bei der Betrachtung der Physiologie des Hörorganes wird gleichzeitig deutlich, dass die menschliche auditive Wahrnehmungsfähigkeit dieser Vielfalt in jedem Fall gerecht zu werden vermag. Dabei sollte jedoch immer bedacht werden, dass auch das Hören und die Interpretation des Gehörten ein individueller Prozess sind, der von einer Reihe an physiologischen und wahrnehmungspsychologischen Faktoren abhängig ist.

Als theoretische Grundlage wurde die mathematische Theorie der Kommunikation nach Shannon (1949) ausgewählt, da diese auf Grund ihrer Differenziertheit geeignet ist die in der vorliegenden Arbeit gestellten Forschungsfragen zu bearbeiten. Denn basierend auf dem Prozess der Signalübertragung werden auch die Eigenschaften von Kommunikator (information source), Rezipient (receiver), Medium (transmitter) und Botschaft (signal) behandelt sowie Kodierung und Dekodierung der Signale und mögliche Störungen (noise) dieses Prozesses miteinbezogen. Diese detaillierte Betrachtungsweise des kommunikativen Prozesses wird der Komplexität der menschlichen sprachlichen Kommunikation am ehesten gerecht.

3. Erregung von Aufmerksamkeit und Vermittlung von Kompetenz

Im folgenden werden auf der Basis vorhandener Forschungsergebnisse prosodische Merkmale von Stimme und Sprechweise identifiziert, die den strategischen kommunikativen Zielen der Werbung, nämlich der Erregung von Aufmerksamkeit sowie der Vermittlung von Kompetenz und dem damit verbundenen Aufbau von Vertrauen und Glaubwürdigkeit, entsprechen.

Es muss allerdings einschränkend darauf hingewiesen werden, dass in der Stimmforschung einfache monokausale Schlüsse nicht möglich sind. Denn die Kombinationsmöglichkeiten einzelner, gleichzeitig oder hintereinander auftretender prosodischer und paralinguistischer Merkmale sind zahllos, sodass es schwierig ist, bestimmte Wirkungen eindeutig einzelnen akustischen Parametern zuzuordnen:

> „Die Schwierigkeiten liegen vor allem darin, dass einerseits die Anzahl der theoretischen Kombinationsmöglichkeiten prosodischer Merkmale enorm groß ist und andererseits eine ebenso große Vielfalt an paralinguistischen Informationen ausdrückbar ist, zu denen so unterschiedliche Phänomene wie Emotionen, Einstellungen, Meinungen, physische Kondition, Geschlecht und Alter des Sprechers gehören.“ (Paeschke 2003:49)

Eine Möglichkeit, trotz dieser Komplexität Forschungsergebnisse zu generieren, ist die Auswahl bestimmter eindeutiger Merkmale, die Messung der akustischen Parameter sowie die auditive Analyse und Beurteilung durch Rezipienten.[41] Wobei grundsätzlich unterschieden werden muss, welche Forschungsziele angestrebt werden: die qualitative, möglichst umfassende Beschreibung von Stimmen, die Zuordnung bestimmter prosodischer Merkmale zu spezifischen Intentionen oder Emotionen, oder die Analyse der Wirkung bestimmter Stimmen und Sprechweisen auf die Zuhörenden. In der vorliegenden Arbeit geht es um die konkreten kommunikativen Ziele der Werbung und die Frage, ob diese als prosodische Merkmale der Sprecherstimmen identifiziert werden können.

Im ersten Kapitel dieses Abschnitts werden nun einleitend die primären kommunikativen Strategien und Ziele der Werbung, Erregung von Aufmerksamkeit und Aufbau von Glaubwürdigkeit durch Vermittlung von Kompetenz, besprochen. Thema des nächstfolgenden Kapitels sind die akustischen Parameter, mit denen die, diesen Zielen entsprechenden, prosodischen Merkmale von Stimmen analysiert werden können. Im dritten Teil dieses Abschnitts werden Forschungsergebnisse besprochen, auf deren Basis dann in der Folge die Signalanalyse und Beurteilung dieser akustischen Parameter erfolgen kann.

---

[41] Enterlein/Bartels/Sendlmeier (2005:10f) weisen in dem Zusammenhang einschränkend darauf hin, dass auditive Analysen immer auch von den beurteilenden Menschen abhängen und dass darüber hinaus die Ergebnisse akustischer Messungen nicht linear sondern nur in wahrnehmungsbezogene Größen umgewandelt sinnvoll interpretiert werden können.

3.1. AUFMERKSAMKEIT UND GLAUBWÜRDIGKEIT ALS STRATEGISCHE ZIELE DER WERBUNG

Damit Werbung als „versuchte Verhaltensbeeinflussung mittels besonderer Kommunikationsmittel“ (Kroeber-Riel/Esch 2011:50) überhaupt wahrgenommen wird, ist es als erstes notwendig, die Aufmerksamkeit sowie die intellektuelle und emotionale Zuwendung potentieller Rezipienten zu gewinnen. Dazu müssen diese „aktiviert“ werden:

> „Als Aktivierung wird ein Zustand vorübergehender oder anhaltender innerer Erregung oder Wachheit bezeichnet, der dazu führt, dass sich die Empfänger einem Reiz zuwenden. Diese Auswirkung der Aktivierung nennt man ‚Kontaktwirkung'. Aktivierung regt außerdem die emotionale und gedankliche Verarbeitung der Reize an. Stark aktivierende Reize werden beispielsweise besser erinnert. Diese Wirkungen kann man als ‚Verstärkerwirkungen' der Aktivierung bezeichnen.
> Die Wirkungen der Aktivierung können in folgender Gesetzmäßigkeit zusammengefasst werden: Je größer die Aktivierungskraft eines Werbemittels ist, umso größer wird seine Chance, unter konkurrierenden Werbemitteln beachtet und genutzt zu werden.“ (Kroeber-Riel/Esch 2011:238)

Um die bestmögliche Aufmerksamkeit und Zuwendung zu erreichen, werden bestimmte Stimuli angewandt: entweder physisch intensive, emotionale oder kognitiv überraschende Reize. Wobei erstere, nämlich große, laute und bunte visuelle und akustische Reize, als ziemlich sichere Methode gelten, um Aktivierung und Zuwendung zum Werbemittel zu erzielen.[42] Im Vergleich zu Radio oder Printmedien bieten Fernsehwerbespots dabei die meisten Gestaltungsmöglichkeiten. Dittmann (1994) weist darauf hin, dass hier über visuelle Mittel wie schnelle Bildfolgen oder Lichteffekte wie auch über akustische Bilder, Lautstärke, Sprechtempo, Stimmqualität, Sprechmelodie sowie durch Verwendung entsprechender Rhythmen eine starke physische Aktivierung ausgelöst werden kann.[43]

Neben Aktivierung als wichtigstem Werkzeug zur Erregung von Aufmerksamkeit ist der Aufbau von Vertrauen, also in der Wahrnehmung und Einschätzung der Zielgruppen vertrauenswürdig zu erscheinen, eines der wesentlichen Ziele der Werbung.[44]

Vertrauenswürdigkeit ist dabei nach Nawratil (2006) neben Kompetenz eine Dimension von Glaubwürdigkeit, wobei es nicht darauf ankommt, ob ein Kommunikator dies wirklich ist, sondern ob es von Seiten der Rezipienten geglaubt wird:

> „Glaubwürdigkeit ist eine Eigenschaft, die eine Quelle nicht von sich aus besitzt, sondern die ihr von ihren Rezipienten zugeschrieben wird. Bei dieser Zuschreibung stützen sich Rezipienten auf eine Reihe von Merkmalen, die in verschiedenen

---

42 Ausführlicher dazu siehe: Kroeber-Riel/Esch 2011:239ff.
43 Siehe dazu: Dittmann (1994) in: Kroeber-Riel/Esch 2011:242.
44 Zu Werbung und Vertrauen siehe ausführlicher: Müller (2009).

> Kommunikationssituationen unterschiedlichen Stellenwert haben. Am häufigsten gründet sich die Einschätzung der Glaubwürdigkeit einer Quelle auf die beiden Dimensionen *Kompetenz* und *Vertrauenswürdigkeit*." (Nawratil 2006:130)

Kompetenz setzt sich dabei nach Nawratil (2006) aus den Faktoren Wissen, Qualifikation, Erfahrung, Leistungen, Intelligenz, sowie dem Innehaben von Führungspositionen zusammen. Vertrauenswürdigkeit steht wiederum „in engem Zusammenhang mit der Ehrlichkeit und Aufrichtigkeit einer Quelle einerseits und ihrer Unparteilichkeit, Unabhängigkeit und Selbstlosigkeit andererseits."(Nawratil 2006:130) Desweiteren können das nonverbale und extra-linguistische Sprechverhalten, sowie auch Sympathie und physische Attraktivität zur Glaubwürdigkeit eines Kommunikators beitragen, wobei das Geschlecht als solches keine Rolle zu spielen scheint.[45]

Kompetenz, Vertrauen und Glaubwürdigkeit stehen auf den ersten Blick - vor allem aus der Sicht kritischer Konsumenten - in Widerspruch zu den Absichten der Werbung, Rezipienten in ihren Einstellungen zu beeinflussen. Doch trotzdem scheinen Werbefachleute davon überzeugt zu sein, diese Diskrepanz überwinden zu können. Müller (2009:99) kommt zu dem Ergebnis, dass „trotz der schwierigen Ausgangsposition der Werbung durch ihre offenkundige und allseits bekannte Parteilichkeit und ihre persuasive Ausrichtung" nach Ansicht von Experten Werbung „keineswegs in einem widersprüchlichen Verhältnis" zu Vertrauen steht und dieses auch von Werbung erreicht werden kann.

Denn in der Werbung geht es eben nicht nur darum mittels Aktivierung auf ein Produkt aufmerksam zu machen, sondern darüber hinaus auch Informationen zu einem Produkt zu vermitteln, die als glaubwürdig anerkannt werden. Das dazu nötige Vertrauen der Rezipienten ist nach Hellmann (2002:4) „unverzichtbar, wenn der Verbraucher nicht mehr weiß, von welcher Qualität die Produkte sind, von deren Beschaffenheit er keine Ahnung mehr hat". Hellmann bezieht sich hier auf Werbung über Massenmedien, bei denen die Kommunikation „anonym und gesichtslos" abläuft. „Man begegnet sich nicht mehr direkt, sondern nur noch auf Distanz, also ohne die Chance, sich noch persönlich zu treffen und kennen zu lernen. Ohne wechselseitige Wahrnehmung und Verständigung fällt aber eine wichtige Voraussetzung des Vertrauens weg." (Hellmann 2002:4) Diesen Mangel an persönlichem Kontakt versucht man in der Werbung durch den Einsatz von Testimonials wie Filmstars oder erfolgreichen Persönlichkeiten aus der Sportszene zu kompensieren. Diese erfüllen nicht nur die Kriterien für Kompetenz sondern haben oft auch hohe Bekanntheits- und Sympathiewerte bei bestimmten Rezipientengruppen und sind daher gut geeignet, den Nachteil der Anonymität in der massenmedialen Kommunikation auszugleichen.[46]

---

[45] Vergleiche dazu: Nawratil (2006:131).
[46] Siehe dazu beispielsweise auch: Schweiger/Schrattenecker (2001:200).

Weitere Werkzeuge, um die Glaubwürdigkeit von Werbung zu erhöhen sind auch die „Belegung besonders glaubwürdiger Medien, Einsatz kompetenter Sprecher, redaktionelle Gestaltung von Anzeigen (Infomercials)“ (Schweiger/Schrattenecker 2001:203).

Eine hohe Glaubwürdigkeit ist aber nicht nur für die von Hellmann (2002) erwähnte Produktinformation, die zum Kauf motivieren soll, ein wichtiger Faktor. Schweiger/Schrattenecker (2001:203) weisen auch auf den interessanten Aspekt hin, dass, je höher die Glaubwürdigkeit der Werbung ist, der Beeinflussungsdruck derselben von den Rezipienten umso weniger wahrgenommen wird.

3.2. AKUSTISCHE PARAMETER

Um Aufmerksamkeits- und Kompetenzsignale im sprachlichen Ausdruck von Stimmen analysieren zu können, bedarf es akustischer Parameter. Basierend auf bisherigen Ergebnissen der Forschung[47] können folgende, für die Analyse dieser prosodischen Merkmale relevante Parameter identifiziert werden: die Sprechgeschwindigkeit, die Grundfrequenz mit den Eigenschaften Mittelwert (oder, wahrnehmungsbezogen ausgedrückt, die Stimmhöhe), Range und Varianz, sowie die Intensität der vokalen Äußerung.

Die Intensität, beziehungsweise die Lautheit als subjektive Wahrnehmung, wird nicht ausführlicher besprochen, da dieser akustische Parameter bei den für diese Arbeit verwendeten Stimmproben der Werbespots von der Lautstärkeregelung der tontechnischen Produktion und Aufzeichnung bestimmt ist und daher bei allen Aufnahmen dieselben Maximum- und Minimumwerte aufweist. Eine vergleichende Analyse ist daher nicht möglich. Der Parameter Intensität ist vor allem im Bereich der natürlichen, spontanen und auch emotional bestimmten Sprache aussagekräftig.

Im folgenden werden die akustischen Parameter Sprechgeschwindigkeit und Grundfrequenz im allgemeinen dargestellt. Im Anschluss daran werden Forschungsergebnisse zur auditiven Perzeption derselben, vor allem hinsichtlich der prosodischen Merkmale Erregung und Kompetenz, besprochen.

3.2.1. SPRECHGESCHWINDIGKEIT

Obwohl sich die Forschung zur Sprechgeschwindigkeit im Lauf des 20. Jahrhunderts zu einem eigenständigen wissenschaftlichen Arbeitsbereich entwickelt hat, gibt es nach wie vor keine eindeutige Definition derselben und auch keine Übereinstimmung darüber, mit welchen Parametern sie gemessen wird.[48]

---

[47] Enterlein/Bartels/Sendlmeier (2005), Paeschke (2003), Schubert/Sendlmeier (2005), Sendlmeier (2005).
[48] Vergleiche dazu ausführlicher: Pfitzinger (2001:129).

Als Indikatoren der Sprechgeschwindigkeit werden im allgemeinen die Dauer beziehungsweise die Länge von Wörtern, Silben und Pausen sowie die Phonrate[49] verwendet. Pfitzinger (2001:124) weist jedoch darauf hin, dass in der deutschen Sprache die Silben- und Phonrate bei kurzen Äußerungsausschnitten bis zu einer Dauer von drei Sekunden mit einem Koeffizienten von unter 0,74 nur mäßig korrelieren und diese beiden Maße daher auch unterschiedliche Informationen widerspiegeln. Diese Verschiedenheit der Phonstrukturen von Wörtern und deren Silben kann mit folgendem Beispiel verdeutlicht werden: „Während das Wort *Banane* drei Silben und sechs Phone, also eine doppelt so hohe Phonrate, aufweist, hat das Wort *schimpfst* nur eine Silbe, aber etwa sieben Phone und damit eine siebenmal höhere Phonrate." (Pfitzinger 2001:124) Beachtet man nun, dass das Verhältnis von Silbe zu Phon 1:2 genauso wie 1:7 sein kann, wird leicht verständlich, dass insbesondere bei kurzen Äußerungen die verwendeten Worte eine unverhältnismäßig große Rolle spielen und eine Vergleichbarkeit dieser Messwerte nicht gegeben ist.

Aus diesem Grund ist es auch sinnvoll die Gesamtlänge der gemessenen Spracheinheiten zu berücksichtigen. Pfitzinger (2001) unterscheidet daher zwischen „globaler" und „lokaler" Sprechgeschwindigkeit. Bei erster wird die Anzahl phonetischer bzw. linguistischer Einheiten eines Redebeitrags, der mindestens einen langen Satz oder mehrere Sätze dauern sollte, durch die akkumulierte Gesamtdauer der Einheiten geteilt. Die Einheiten (Phone, Silben oder Wörter) werden pro Sekunde angegeben. Bei der lokalen Sprechgeschwindigkeit wird hingegen in gleichmäßigen Abständen von z. B. 20 ms je ein Messwert berechnet, „indem man mit Hilfe einer um den jeweiligen Messpunkt zentrierten Fensterfunktion einen Signalausschnitt von z. B. 500 ms Dauer extrahiert und aus ihm dann die Einheiten pro Sekunde ermittelt". Daraus ergibt sich dann „alle 20 ms ein lokaler Sprechgeschwindigkeitswert und damit im Ganzen eine synchron zum Signal verlaufende Sprechgeschwindigkeitskurve, die bei langsamen Äußerungsteilen einen niedrigeren und bei schnellen einen entsprechend höheren Wert aufweist". (Pfitzinger 2001:139)

Insbesondere bei der Messung globaler Sprechgeschwindigkeit sind jedoch auch die Pausen, also jene Zeitabschnitte, in denen keine Laute produziert werden, zu berücksichtigen. Während des Sprechens sind Pausen einerseits zum Einatmen biologisch notwendig, andererseits können sie aber auch eingesetzt werden um das Gesagte zeitlich zu strukturieren und um bestimmte Inhalte zu betonen.[50] Bei der Berechnung der Sprechgeschwindigkeit sollte nach Paeschke (2003:28) die Pausendauer jedoch nicht miteinbezogen sondern als eigener Parameter angesehen werden. Winkler (2008) misst beispielsweise bei spontanem Sprechen nicht nur die Anzahl sondern auch den prozentuellen Anteil der Pausen an der Länge des Gesamtsignals. Dieser Anteil korreliert nach

---

[49] Ein Phonem wird in der Phonologie als die Norm eines Sprachlautes verstanden. Dieses kann, muss aber sehr oft auch nicht, mit der entsprechenden Silbe übereinstimmen. Daher wird stattdessen der Begriff Phon verwendet. Ein Beispiel für die Notwendigkeit dieser Unterscheidung sind die Wörter „Frauchen" und „rauchen". Beide bestehen aus je zwei Silben, doch muss das „F" bei „Frauchen" als eigener Laut, das heißt, Phon kenntlich gemacht werden, da nur so die unterschiedliche Aussprache und Bedeutung erkennbar wird. Siehe dazu ausführlicher: Trojan (1975: 26ff), Pfitzinger (2001:166f).

[50] Siehe dazu ausführlicher: Schubert/Sendlmeier (2005:19f).

Schubert/Sendlmeier (2005) in der Perzeption allerdings wieder mit der Sprechgeschwindigkeit, die als umso langsamer wahrgenommen wird je länger die Pausendauer ist.

Für die vorliegende Untersuchung von Werbespots erscheint es sinnvoll, die globale Sprechgeschwindigkeit unter Berücksichtigung der Pausen zu messen. Nachdem die in diesem Themenfeld relevanten Forschungsergebnisse Silben pro Sekunde angeben, wird ebenfalls die Silbenrate als Indikator herangezogen um die jeweiligen Werte vergleichbar zu machen.

3.2.2. STIMMGRUNDFREQUENZ

Die Beschreibung und Berechnung der Grundfrequenz eines Tones ist umso komplexer je vielfältiger derselbe ist. Ein reiner Ton, auch Sinuston genannt, besteht aus einer einzigen Schallwelle, deren momentaner Schalldruck einer einfachen sinusoidalen Funktion über die Zeit folgt. Sinustöne kommen in der Natur nicht vor, können jedoch künstlich erzeugt werden. Sie werden beispielsweise für Stimmgeräte in der Musik verwendet. Sprache, Klänge und Geräusche bestehen hingegen aus einer Vielzahl an Schallwellen mit unterschiedlichen Schwingungsfrequenzen. Bei Geräuschen sind diese Schwingungen unregelmäßig und statistisch zufällig, bei (harmonischen) Klängen stehen sie zueinander in einem ganzzahligen Verhältnis.[51] Im Zusammenhang der vorliegenden Fragestellungen werden jedoch keine Teiltöne oder Formanten (stark ausgeprägte Frequenzbereiche) berücksichtigt sondern alleine der Verlauf und die Eigenschaften der Grundfrequenz analysiert.

Bei der menschlichen Stimme entspricht die Grundfrequenz dem Reziprok der Periodendauer einer Stimmlippenschwingung. Dies gilt für einen einzelnen Laut. Während des Sprechens, der Artikulation, ändert sich jedoch mit den unterschiedlichen Lauten und Vokalen sowie der Sprachmelodie die Grundfrequenz permanent. Terhardt (1998:183) spricht diesbezüglich von einem höchst dynamischen Vorgang, bei dem die aufeinanderfolgenden Laute weder bezüglich der Artikulation noch der Struktur des Schallsignals scharf voneinander abgegrenzt sind.

Es ist demzufolge nicht möglich, einen absoluten Wert der Grundfrequenz einer sprachlichen Äußerung festzulegen. Man kann jedoch den Verlauf der Grundfrequenz analysieren und den Mittelwert der in bestimmten zeitlichen Abständen gemessenen Grundfrequenzwerte bestimmen. So kann ein Richtwert über die mittlere Stimmhöhe einer sprachlichen Äußerung berechnet werden.

Der Mittelwert alleine sagt jedoch nichts über die Verteilung der Daten beziehungsweise über die Charakteristika der Grundfrequenz der sprachlichen Äußerung aus. Dazu bedarf es weiterer

---

[51] In dem Zusammenhang wird auch von Teiltönen (bei inharmonischen Klängen) oder von Harmonischen (Grundtönen und Obertönen) gesprochen. Sie sind ein bestimmender Faktor für die spezielle Charakteristik von Klängen (Klangfarben).

Parameter: Quantile als verteilungsannahmefreie Lagemaße, sowie die Maße der Streuung Standardabweichung (Quadratwurzel der Varianz) und Range (Spannweite).

Mit Range wird die Differenz des höchsten und des niedrigsten gemessenen Wert bezeichnet. Daraus ist die Spannweite der Grundfrequenz ablesbar.

Bei der Messung der Grundfrequenz können auch sogenannte Ausreißer[52] nach oben oder unten vorkommen, die zum Beispiel durch bestimmte Laute wie „s“ oder „n“ erzeugt werden und die die tatsächliche Charakteristik verzerren. Man verwendet daher den Median und andere Quantile, die, verglichen mit dem arithmetischen Mittel, statistisch robuster sind, also den Einfluss von Ausreißern reduzieren.

Mit der Standardabweichung wird gemessen, wie stark die Daten um ihren Mittelwert streuen. Als Varianz bezeichnet man die Abweichungsquadratsumme dividiert durch n, wobei n die Anzahl der Messungen ist. Somit werden alle Abweichungen (positive wie auch negative) auf ein positives Vorzeichen umgerechnet. Ist der Varianzwert klein, streuen die Daten eng um den Mittelwert, ist er groß, ist eine breite Streuung vorhanden. Das Streuungsmaß der Quantile (Anteilswerte) ist der Interquantilsabstand, der ebenfalls verteilungsannahmenfrei ist.[53]

Zusammenfassend und wahrnehmungsbezogen ausgedrückt kann gesagt werden, dass der Mittelwert und der Median der Grundfrequenz eine Vorstellung über die Tonhöhe der stimmlichen Äußerung geben. Die Streuung, der Range und die Interquantilsabstände der Grundfrequenz sind weitere wichtige Parameter, die die Lebendigkeit der Sprechweise und sprachlichen Äußerung charakterisieren.

3.3. Erregung und Kompetenz als prosodische Merkmale des stimmlichen Ausdrucks

Die weiter oben erwähnten lauten und intensiven akustischen Reize zur Aktivierung finden in der Stimmforschung eine Entsprechung in der Dimension Erregung. Deren Parameter sind nach Sendlmeier (2005:10) die Sprechgeschwindigkeit, die Grundfrequenz und Intensität. In der auditiven Perzeption entspricht die Grundfrequenz der Wahrnehmung der Tonhöhe, die Intensität der Empfindung von Lautheit. Keines dieser Merkmale bleibt jedoch über den Verlauf einer Äußerung konstant. Insbesondere die Varianz der Grundfrequenz ist dabei nicht nur ein Merkmal für monotone (geringe Varianz) oder lebendige (große Varianz) Sprechweise, sondern auch für den

---

[52] Als Ausreißer werden Beobachtungswerte bezeichnet, die von den übrigen in ihrer Größenordnung stark abweichen. Sie beeinflussen auch die Ergebnisse bei der Berechnung der Standardabweichungen beziehungsweise Varianz. (Vgl. dazu: Stier 1999:259)

[53] Grundlagen und Formeln zur Berechnung der statistischen Parameter sind nachzulesen bei Fahrmeir et al. (2011). Für die Analyse der ausgewählten Fernsehwerbespots wird das von Paul Boersma und David Weenink entwickelte Sprachanalyseprogramm Praat verwendet, das diese Parameter auf der Basis spezieller Algorithmen automatisiert berechnet.

emotionalen Ausdruck und den Grad der Erregung (eine große Varianz bedeutet ein hohes Maß an Erregung).[54]

Eine schnelle Sprechgeschwindigkeit von etwa sieben Silben pro Sekunde, erhöhte Stimmlage sowie eine große Varianz und ein großer Range sind nach Enterlein/Bartels/Sendlmeier (2005) auch Merkmale von positiv wertenden Äußerungen. Darüber hinaus sind auch die Emotionen Freude und Ärger durch ein hohes Maß an Erregung charakterisiert:

> „Freude ist eine der Emotionen, die durch sehr hohe Erregung gekennzeichnet sind. Dies schlägt sich insbesondere in der hohen Stimmlage und dem stark erhöhten Stimmumfang nieder. Die mittlere Grundfrequenz liegt 10 bis 11 Halbtöne[55] höher als bei neutraler Sprechweise, der Grundfrequenzrange ist etwa 7 Halbtöne größer. Diese Werte entsprechen den Werten ärgerlicher Äußerungen und sind wesentlich größer als die Werte der anderen Emotionen." (Paeschke 2003:222)

Bei der Varianz der Grundfrequenz, die der dritte starke Indikator für emotionale Erregung ist, sind nach dieser Untersuchung von Paeschke (2003) die Messergebnisse für Freude noch um etwa 10% stärker als diejenigen für Ärger. Bei der Lautheit hingegen sind die Werte für Ärger wieder etwas höher als bei Freude.[56] Die Werte für alle anderen gemessenen Emotionen (Angst, Trauer, Ekel, Langeweile) liegen wesentlich unter diesem hohen Erregungsgrad. Bei der Sprechgeschwindigkeit hingegen zeigte sich, dass die Werte für die Emotion Angst mit 6,75 Silben pro Sekunde deutlich höher lagen als für Freude oder Ärger. Paeschke interpretiert dieses Ergebnis so, dass bei Angst „der größte Teil der Erregung in eine stark erhöhte Sprechgeschwindigkeit umgesetzt" wird (Paeschke 2003:228).

Die Sprechgeschwindigkeit ist aber nicht nur ein Indikator für verschiedene Formen der emotionalen Erregung. Eine schnelle Redeweise wird in der Perzeption auch mit Kompetenz in Verbindung gebracht, da dann, im Unterschied zu einem eher langsameren Tempo, eher davon ausgegangen wird, dass die sprechende Person weiß, wovon sie redet.[57] Auch eine deutliche Variation der Tonhöhe wird im Vergleich zu monotoner Sprechweise als Zeichen für Kompetenz und Selbstbewusstsein gewertet.[58]

---

[54] Vergleiche dazu ausführlicher: Enterlein/Bartels/Sendlmeier (2005).

[55] Ein Halbton ist das in unserer westlichen Musikkultur kleinste gebräuchliche Intervall zwischen zwei Tönen. Eine Oktave (siehe Fußnote 29) kann in zwölf Halbtonschritte unterteilt werden. Nach diesen Untersuchungsergebnissen von Paeschke (2003) wird bei Freude also fast eine Oktave höher gesprochen als bei neutraler Sprechweise. Dies entspricht annähernd einer Verdoppelung der Frequenz.

[56] Paeschke untersucht die Unterschiede der prosodischen Merkmale von Emotionen wesentlich umfassender als im vorliegenden Zusammenhang angesprochen werden kann. Bei detaillierter Betrachtung ist daher Ärger auch deutlicher unterscheidbar von Freude als aufgrund der hier angesprochenen Parameter zu vermuten wäre. Insbesondere die intensive und rasche Art der Betonung von Silben, „was beim Zuhören den unangenehmen Eindruck erzeugen kann, dass man mit Worten geschlagen wird" (Paeschke 2003:226), ist für den Ausdruck von Ärger charakteristisch.

[57] Vergleiche dazu: Felser (2007:324), Nawratil (2006:84f).

[58] Siehe dazu: Eckert/Laver (1994:34).

Ein bestimmender Faktor bei der Vermittlung von Glaubwürdigkeit scheint jedoch die durchschnittliche Tonhöhe von Stimmen zu sein. Eckert/Laver kommen in ihren Untersuchungen zu dem Ergebnis, dass bei den Versuchspersonen die Tendenz bestand, „tiefe Männerstimmen als angenehm und als Ausdruck von Souveränität, Vertrauenswürdigkeit, Autorität etc. einzuschätzen“ (Eckert/Laver 1994:37). Auch Schubert/Sendlmeier stellen fest, dass in den westlichen Kulturen „eine starke Präferenz für tiefe Stimmen zu beobachten“ ist. Denn in der Wahrnehmung der Rezipienten signalisieren diese neben Ruhe auch Kompetenz, Glaubwürdigkeit und Dominanz. (Schubert/Sendlmeier 2005:15)

Allerdings weisen Eckert/Laver darauf hin, dass diese Beurteilungen nur für Stimmen gelten, die in ihrer natürlichen Indifferenzlage, das heißt, in der Tonhöhe, die den physiologischen Voraussetzungen am besten entspricht, verwendet werden. Je weiter sich die Stimme jedoch permanent und gewohnheitsmäßig entweder nach oben oder unten von diesem optimalen Frequenzbereich entfernt, „desto unnatürlicher wird sie vom Hörer empfunden und entsprechend negativ eingeschätzt“ (Eckert/Laver 1994:39). Wenn hingegen „die durchschnittliche Stimmhöhe in der Indifferenzlage des Sprechers bzw. der Sprecherin liegt, empfinden die meisten Hörer das als natürlich und angenehm“ (Eckert/Laver 1994:162).

Wenn auch bei Männern „die natürlich klingenden tieferen Sprechstimmen oft positiv und als Ausdruck von Kompetenz, manchmal sogar von Autorität bewertet“ werden (Eckert/Laver 1994:163), so gilt dies jedoch nicht grundsätzlich für jede tiefe Stimme, da selbstverständlich auch weitere Merkmale eine Rolle spielen: „Jeder Sprecher hat noch eine Reihe von anderen Stimmeigenschaften, die mitbeurteilt werden. Eine zwar tiefe, aber gequetschte und nicht voll klingende Stimme wurde weniger positiv oder gar negativ beurteilt.“ (Eckert/Laver 1994:37)

Abschließend kann somit festgestellt werden, dass tiefe Männerstimmen, die in der natürlichen Indifferenzlage gesprochen werden und die keine weiteren Merkmale haben, die einen vollen tiefen Klang stören könnten, im allgemeinen als angenehm, souverän und vertrauenswürdig beurteilt werden. Eckert/Laver bemerken in diesem Zusammenhang, dass es daher „nicht verwunderlich“ sei, dass „tiefe Männerstimmen sehr häufig am Ende eines Reklamespots für die zusammenfassende Bewertung des Produktes eingesetzt werden“ (Eckert/Laver 1994:37).

3.4. ZUSAMMENFASSUNG

Werbung verfolgt erstens das grundlegende Ziel, Aufmerksamkeit zu erregen und dadurch Zuwendung zu erreichen, zweitens soll Kompetenz ausgestrahlt und dadurch Glaubwürdigkeit vermittelt werden. Thema der vorliegenden Arbeit ist die Frage, ob diese Intentionen bei den in Werbespots verwendeten Stimmen als prosodische Merkmale Ausdruck finden und über akustische Parameter gemessen werden können.

Die für diese Messungen relevanten Parameter sind die Sprechgeschwindigkeit und die Grundfrequenz der Stimme mit ihren Eigenschaften Mittelwert, Range und Varianz.

Aufmerksamkeit und Aktivierung durch starke Reize zu erzeugen entspricht in der Stimmforschung der Dimension der Erregung. Prosodische Merkmale einer starken Erregung sind erhöhte Stimmlage, hohe Varianz, großer Range, schnelle Sprechgeschwindigkeit und auch erhöhte Intensität. Vor allem die Emotion Freude wird durch diese prosodischen Merkmale der starken Erregung ausgedrückt. Nur bei Ärger ist die Intensität höher, bei Angst hingegen die Sprechgeschwindigkeit.

Eine schnelle Sprechgeschwindigkeit und erhöhte Varianz sind auch Merkmale, die in der auditiven Perzeption mit Kompetenz in Verbindung gebracht werden. Der bestimmende Faktor und wesentliche Unterschied zum Ausdruck von Erregung ist jedoch die Höhe der Grundfrequenz. In unserer Sprachkultur wird Kompetenz vor allem tiefen Männerstimmen zugeordnet, die auch mit Vertrauenswürdigkeit und Souveränität in Verbindung gebracht werden.

Diese Ergebnisse aus der Stimmforschung legen den Schluss nahe, dass Stimmen in der Fernsehwerbung nur eines der beiden kommunikativen Ziele der Werbung ausdrücken können: entweder Aufmerksamkeit und Aktivierung oder Kompetenz. Mit der folgenden Signalanalyse von Fernsehwerbespots soll nun untersucht werden, ob prosodische Merkmale der darin eingesetzten Stimmen und Sprechweisen identifiziert werden können, die einem dieser Ziele zuordenbar sind.

4. SPRECHGESCHWINDIGKEIT UND STIMMGRUNDFREQUENZANALYSE VON FERNSEHWERBESPOTS

Im folgenden Kapitel wird einleitend der Prozess der Auswahl von zur Stimmanalyse geeigneten Fernsehwerbespots beschrieben. Danach werden die Berechnungen der akustischen Parameter Sprechgeschwindigkeit und Grundfrequenz der verwendeten Stimmen zuerst einzeln dargestellt, im Anschluss daran werden sie gemeinsam und vergleichend analysiert und interpretiert.

Zur Messung der akustischen Parameter wird das Stimmanalyseprogramm Praat verwendet, das von Paul Boersma und David Weenink am Institute of Phonetics Sciences an der Universität Amsterdam entwickelt und erstmalig 2003 veröffentlicht wurde. Seitdem wurde das Programm ständig erweitert und verbessert. Für die vorliegende Arbeit wurde die Version von 2015 verwendet. Praat ist als Freeware erhältlich (http://www.praat.org). Es sind dafür keine speziellen Systemvoraussetzungen am Rechner notwendig, allerdings sind zumindest ein 64 Bit-Betriebssystem und 2 GB Arbeitsspeicher erforderlich um die Berechnungen problemlos durchführen zu können.

Praat wurde als Rechenprogramm ausgewählt, da es im Unterschied zu anderen, die vor allem für Musik eingesetzt werden, für die akustische Analyse von menschlichen Stimmen und gesprochener Sprache optimiert ist. Desweiteren ermöglicht dieses Programm sehr detaillierte Berechnungen der Signalparameter.

In den vorangegangenen Kapiteln wurde mehrmals auf die Komplexität der akustischen Analyse von gesprochener Sprache hingewiesen. Weil bei vielen Parametern technisch keine absoluten Messergebnisse erzeugt werden können, werden stattdessen auf Basis von Algorithmen Schätzwerte („Estimated Parameter“) errechnet.[59] Diese heutigen Methoden der Signalanalyse sind jedoch erst möglich, seit leistungsstarke, auf digitaler Technologie basierende Rechner mit großen Arbeitsspeichern verfügbar sind.

Im allgemeinen sind Einzelergebnisse bestimmter Signalparameter nicht sinnvoll interpretierbar, sondern benötigen den Bezug zu Vergleichs- oder Richtgrößen. In der folgenden Analyse werden daher immer auch Rangreihen der Messergebnisse erstellt, um so relationale Aussagen zur Bewertung derselben treffen zu können.

Die Werbespots wurden auf Videodateien aufgezeichnet, dann für jeden Spot die Tonspur isoliert und von einer Windows Media Datei in eine WAV-Audiodatei konvertiert. Diese Audiodaten wurden mit Hilfe von Praat analysiert. Die mit Praat berechneten Amplitudenverläufe sowie die detaillierten Messwerte der Stimmgrundfrequenzen sind in Anhang 1 und 2 dargestellt.

---

[59] Die bei Praat verwendeten Algorithmen können im dazugehörigen Handbuch nachgelesen werden. Siehe dazu: http://www.praat.org (15.8.2015).

4.1. AUSWAHL DER FERNSEHWERBESPOTS

Ziel bei der Auswahl der Fernsehwerbespots ist es, sprachlich passende Spots zu finden, die mit Focus auf Stimme und Sprechweise analysiert und auch verglichen werden können.

Das erste Kriterium dabei ist die alleinige Verwendung einer einzelnen Stimme als akustisches Mittel, das heißt, ohne Musik, möglichst ohne zusätzliche Geräusche und ohne Dialoge. Bei mehreren Stimmen würden aufgrund der Komplexität und Inhomogenität der Stimmen eindeutige Ergebnisse oder Vergleiche von verschiedenen Spots nicht möglich sein. Musik oder markante Geräusche, die gleichzeitig oder auch abwechselnd mit Stimmen eingesetzt werden, verändern hingegen nicht nur die Berechnungen und die Analyse der Tonsignale, auch der Höreindruck ist ein anderer. Es kann auch hypothetisch davon ausgegangen werden, dass die Sprechstimmen anders eingesetzt werden, wenn sie Bestandteil und nicht alleiniges Mittel des akustischen Ausdrucks sind. Diese Thematik könnte zu weiteren Forschungsfragen führen. Für die vorliegende Arbeit mussten jedoch Spots gefunden werden, in denen die ganze Werbebotschaft von nur einer Stimme gesprochen wird. Dabei wurde auch darauf geachtet, dass die Sprache der Spots Hochdeutsch ist und weder ein Dialekt noch ein markanter Akzent verwendet werden. Da sich bei Dialekt und Akzent die Stimmen anders verhalten, wäre auch hier eine vergleichende Analyse der Spots nicht möglich.

Im Zeitraum zwischen Juli bis Dezember 2007, Februar bis August 2008, sowie Februar bis September 2012 wurden immer wieder Stichproben von Werbesendungen, die in ORF1, ORF2 und ATV gesendet wurden, auf Windows Media Dateien aufgezeichnet. Aus diesen Stichproben konnten insgesamt 13 Spots ausgesucht werden, die folgenden Kriterien entsprechen:

- Die Werbebotschaft wird von einer Stimme alleine gesprochen.
- Es gibt keine durchgehende Musik im Hintergrund.
- Die Sprache der Spots ist Hochdeutsch, kein Dialekt, kein markanter Akzent.

Diese 13 Spots sind in Tabelle 1 aufgelistet. Die Reihung erfolgt dabei nach der Sprechdauer. In der ersten Spalte der Tabelle wird der Produktname gelistet, dann die Gesamtdauer der Aufnahme sowie die Sprechdauer in Sekunden mit zwei Nachkommastellen[60]. In der nächsten Spalte wird angeführt, ob die Stimme von einer Frau oder von einem Mann ist. Dabei stellt der Spot Möbelix einen Sonderfall dar, da hier eine Männerstimme wie in einem Comicfilm tontechnisch verändert wurde.

[60] Praat arbeitet mit sechs Nachkommastellen. Diese Genauigkeit wird für das Bearbeiten der Tonspuren benötigt. Für die vorliegenden Berechnungen ist dies jedoch nicht notwendig und wird daher auf zwei Nachkommastellen gerundet.

Werbespot	Gesamtdauer in Sekunden	Sprechdauer in Sekunden	Stimme	Hintergrund-geräusch	Kommentar
Leiner	6	3,55	Frau	ja	leichte Hintergrundgeräusche, wie auf einer Straße, entferntes Hupen
Möbelix	6	4,62	Mann*	nein	* Stimme tontechnisch verändert
Nestea	10	9,25	Mann	ja	permanente Naturgeräusche, Wind, Regen, Donnergrollen
Alpecin	10	9,40	Mann	nein	in der ersten Sekunde gibt es gleichzeitig mit der Stimme ein akustisches Signal
Neuroth	14	12,70	Mann	nein	
VW	16	13,39	Mann	ja	permanente Hintergrundgeräusche, Vogelzwitschern, Automotor, Autotür
WC-Ente	16	14,96	Mann	nein	
Canesten Glutrimazol	16	15,68	Frau	ja	ab der 11. Sekunde wird ca. vier Sekunden lang Musik im Hintergrund gespielt; die letzten vier Sekunden des Spots mit dem Standardtext: „Über Wirkungen und Nebenwirkungen informieren sie Arzt oder Apotheker“ wurden weggeschnitten, da sie von einer anderen Stimme gesprochen und für die Analyse nicht relevant sind.
Innocent	20	18,24	Mann	ja	permanente Naturgeräusche, Vogelzwitschern, Rauschen
CafeHaag	21	18,64	Frau	ja	permanentes Geschirrklappern, Geräusche wie im Kaffeehaus, Gemurmel
Plantur	21	19,41	Frau	nein	
Sensodyne	21	19,84	Frau	nein	
Canesten Bifonazol	26	24,79	Mann	nein	

Tabelle 1: Überblick über die analysierten Werbespots

Hintergrundgeräusche sind hinsichtlich der akustischen Analyse von Relevanz. Denn bei den Berechnungen der Signale muss die Stimme von anderen Geräuschen getrennt werden können.

Diese Unterscheidung erfolgt beim menschlichen Hörorgan selbstverständlich, wie weiter oben ausgeführt wurde, bei der Signalanalyse ist dies jedoch nur eingeschränkt möglich. Es wird daher auch angegeben, ob Hintergrundgeräusche vorhanden sind. Diese werden in der letzten Spalte kommentiert.

In einem nächsten Arbeitsschritt wurden mit Praat von allen Spots Grafiken erstellt, in denen der Amplitudenverlauf der Tonaufnahmen in einer Wave-Anzeige visualisiert ist. Diese Grafiken sind zusammen mit den in den Spots gesprochenen Texten sowie mit den Berechnungen der Pausen und Sprechgeschwindigkeiten in Anhang 1 aufgelistet.

4.2. Berechnung der Sprechgeschwindigkeit

Zur Berechnung der Sprechgeschwindigkeit wird die Anzahl der Silben des Textes durch die Sprechdauer in Sekunden dividiert. Pausen ab einer Länge von 0,25 Sekunden werden dabei ausgewiesen und ebenfalls berechnet. Dies entspricht nach Haselow (2015:105) der Untergrenze von kurzen Pausen, die mit einer Länge von 0,25 bis 0,75 Sekunden definiert sind.

In Anhang 1 werden die mit Hilfe von Praat erzeugten Grafiken der Amplitudenverläufe der Spots abgebildet. Desweiteren werden die gesprochenen Texte dokumentiert, sowie die Gesamtdauer der Aufnahme, die Sprechdauer in Sekunden mit zwei Nachkommastellen (in Klammer gesetzt ist die Anfangs- und Endzeit der Sprechdauer), die Differenz der Sprechdauer in Prozent und das jeweilige Ergebnis, gerundet auf zwei Nachkommastellen, für die Sprechgeschwindigkeit für jeden einzelnen Spot berechnet und aufgelistet. Pausen werden mit ihrer Länge in Sekunden (sec), auf zwei Nachkommastellen gerundet, an den entsprechenden Stellen der Texte in eckige Klammern gesetzt. Um einen möglichen Einfluss der Pausen auf die Sprechgeschwindigkeit deutlich machen zu können, wird diese zweimal berechnet: einmal mit der Gesamtdauer (global), einmal unter Abzug der Pausenzeiten. Angeführt ist ebenfalls die Differenz dieser zwei Ergebnisse.

In der Beschreibung zu diesen Berechnungen wird auch der Einfluss der Hintergrundgeräusche deutlich gemacht. Denn bei der Signalanalyse mit Praat mussten für einzelne Spots die Einstellungen für die Grenzwerte des Schalldruckpegels (Dezibel, dB) an die Hintergrund-geräusche angepasst werden, um diese von den Stimmen zu trennen und Sprechpausen erkennen zu können. Um diese spezifischen Werte im Rechenprogramm korrekt angeben zu können, sind akustische Kontrollen unerlässlich.

In Tabelle 2 werden die Ergebnisse dieser eigenen Berechnungen zusammengefasst. Die Reihung der Spots erfolgt hier nach der globalen Sprechgeschwindigkeit (fett gedruckt, in der fünften Spalte), beginnend beim langsamsten Spot Leiner bis zum schnellsten Spot Möbelix. In der zweiten Spalte wird gesamte Sprechdauer angeführt, in der dritten die Sprechdauer minus der Pausen, sofern vorhanden, aufgelistet. In der vierten Spalte wird die Differenz der Sprechdauer bei Pausen

in Prozent angegeben. In der fünften Spalte wird die globale Sprechgeschwindigkeit (Silben pro Sekunde) angegeben, bei der Pausen nicht berücksichtigt sind. In der sechsten Spalte wird die Sprechgeschwindigkeit unter Berücksichtigung der Pausen (Silben pro Sekunde) angegeben und in der siebenten die Differenz der Silben pro Sekunde bei Pausen.

Bei vier Spots, Nestea, WC-Ente, Alpecin und Möbelix, wurden keine Pausen über 0,25 Sekunden Länge festgestellt.

Die Veränderung der Sprechdauer durch Pausen bewegt sich zwischen rund 4% (Sensodyne) und fast 25% (Cafe Haag). Dementsprechend ist auch die Differenz der gesprochenen Silben pro Sekunde bei dem Spot Cafe Haag mit 1,15 Silben wesentlich höher als bei den anderen Spots, die mit Werten von 0,24 (Sensodyne) bis 0,79 (VW) alle deutlich unter einer Silbe liegen.

Spot	**Dauer gesamt in Sekunden**	**Dauer min. Pausen in Sekunden**	**Dauer Differenz %**	**Silben /Sek. global**	**Silben /Sek. min. Pausen**	**Silben /Sek. Differenz**	**Stimme**
Leiner	3,55	2,91	18,03	**3,38**	4,12	0,74	Frau
Neuroth	12,70	11,03	13,15	**3,39**	3,90	0,51	Mann
Cafe Haag	18,64	14,07	24,52	**3,54**	4,69	1,15	Frau
Innocent	18,24	16,80	7,89	**3,89**	4,23	0,34	Mann
Can. Bi.	24,79	21,45	13,47	**4,03**	4,66	0,63	Mann
Nestea	9,25	-		**4,22**	-		Mann
VW	13,39	11,39	14,94	**4,48**	5,27	0,79	Mann
WC-Ente	14,96	-		**4,48**	-		Mann
Can. Gl.	15,68	14,89	5,04	**4,78**	5,04	0,26	Frau
Plantur	19,41	16,70	13,96	**4,79**	5,57	0,78	Frau
Alpecin	9,40	-		**5,21**	-		Mann
Sensodyne	19,84	19,01	4,18	**5,34**	5,58	0,24	Frau
Möbelix	4,62	-		**5,63**	-		Mann*

Tabelle 2: Sprechgeschwindigkeit der Werbespots

Spot	Silben /Sek. global	Rang	Silben /Sek. min. Pausen	Rang	Rang Differenz	Differenz quadriert
Leiner	**3,38**	1	4,12	2	-1	1
Neuroth	**3,39**	2	3,90	1	1	1
Cafe Haag	**3,54**	3	4,69	7	-4	16
Innocent	**3,89**	4	4,23	4	0	0
Can. Bi.	**4,03**	5	4,66	6	-1	1
Nestea	**4,22**	6	-	3	3	9
VW	**4,48**	7,5	5,27	10	-2,5	6,25
WC-Ente	**4,48**	7,5	-	5	2,5	6,25
Can. Gl.	**4,78**	9	5,04	8	1	1
Plantur	**4,79**	10	5,57	11	-1	1
Alpecin	**5,21**	11	-	9	2	4
Sensodyne	**5,34**	12	5,58	12	0	0
Möbelix	**5,63**	13	-	13	0	0
					$\sum$ 0,0	$\sum$ 46,5

Tabelle 3: Sprechgeschwindigkeit mit und ohne Berücksichtigung der Pausen

Es stellt sich die Frage, ob bei einer Berücksichtigung der Pausen die Differenzen von unter einer Silbe pro Sekunde genügend Aussagekraft für die vorliegenden Forschungsfragen haben. Einen ersten Hinweis dazu kann ein Vergleich der Spots mit und ohne Berücksichtigung der Sprechpausen geben. (Siehe dazu Tabelle 3.) Dazu werden zwei Rangreihen der Sprechgeschwindigkeiten der Spots einmal mit und einmal ohne Pausen (Silben pro Sekunde) erstellt, die Differenzen der Ränge ermittelt und die quadrierten Differenzen summiert. In weiterer Folge wird der Rangkorrelationskoeffizient nach Spearman (r_s) berechnet.

$$r_{sp} = 1 - \{6\sum d_i^2 /[n(n^2-1)]\}$$

$$r_{sp} = 1 -\{6*46,5/[13*(169-1)]\} = 0,87225$$

Es erstaunt nicht, dass mit einem Koeffizienten von 0,87 ein starker Zusammenhang festgestellt werden kann, da die Daten der zwei Rangreihen voneinander abhängig sind und die meisten Ränge nahe beieinander liegen sowie drei davon (Innocent, Sensodyne, Möbelix) gleich geblieben sind. Auf dieser Basis kann durchaus gerechtfertigt argumentiert werden, dass für die vorliegenden Forschungsfragen eine Berücksichtigung der Pausenzeiten nicht notwendig ist. Dies wird auch unterstützt durch die schon erwähnten Forschungsergebnisse von Schubert/Sendlmeier (2005), die darauf hinweisen, dass in der akustischen Perzeption die Sprechgeschwindigkeit als umso langsamer wahrgenommen wird je länger die Pausendauer ist. Für die weiteren Analysen werden daher die niedrigeren Werte der globalen Sprechgeschwindigkeiten herangezogen.

In Spalte acht von Tabelle 2 wird auch das Geschlecht der Sprechstimme angegeben. Hier ist deutlich erkennbar, dass kein Zusammenhang zwischen männlicher oder weiblicher Sprechstimme und Sprechgeschwindigkeit hergestellt werden kann. Beide Stimmen sind jeweils bei niedrigen, mittleren und höheren Werten vertreten. Beachtenswert ist dabei, dass die Comic-Stimme des Spots Möbelix die höchste Sprechgeschwindigkeit hat.

Zusammenfassend kann festgestellt werden, dass die Sprechgeschwindigkeiten der Spots Werte von 3,38 (Leiner) bis 5,63 (Möbelix) Silben pro Sekunde aufweisen. Der Mittelwert der globalen Sprechgeschwindigkeit aller 13 Spots liegt bei 4,40 Silben pro Sekunde, was annähernd den zwei Spots VW und WC-Ente (verbundener Rang 7,5 von 13) entspricht. Eine Gewichtung Richtung niedrigerer oder höherer Werte ist somit nicht feststellbar.

In Kapitel 4.4 werden diese Ergebnisse interpretiert und auch mit den Ergebnissen der Analyse der Grundfrequenz zusammengeführt.

4.3. Beschreibung der Stimmgrundfrequenz

Bei den Amplitudenverläufen, die in Anhang 1 abgebildet sind, wurde bereits der Störfaktor der Hintergrundgeräusche deutlich. Um die Sprechpausen eindeutig identifizieren zu können, mussten die Messwerte für die Schalldruckpegel jeweils individuell angepasst werden. Zur Analyse der Grundfrequenzen der Sprechstimmen sind diese Spots somit nicht geeignet, da die Frequenzbereiche der Stimmen durch diejenigen der Hintergrundgeräusche überlagert und nicht trennbar sind. Die Spots Leiner, Cafe Haag, Innocent, Nestea und VW werden daher nicht weiter bearbeitet.

Im folgenden wird nun die Grundfrequenz der verbleibenden acht Werbespots zuerst einzeln und dann vergleichend beschrieben. Die Reihung in der Darstellung erfolgt aufsteigend nach der Länge der Spots. Eine detaillierte Auflistung aller mit dem Programm Praat errechneten Werte findet sich in Anhang 2. Die Messung erfolgte im Frequenzbereich zwischen 30 – 600 Hz. In diesem Analysefenster werden Teiltöne und Formanten nicht erfasst. Es ist allerdings groß genug, um spezielle tiefe und hohe Signale, die insbesondere durch Artikulation entstehen, sehen zu können.

Die qualitative Bewertung und Interpretation hinsichtlich der Aufmerksamkeits- und Kompetenzsignale erfolgt dann im fünften Abschnitt.

Bei den von Praat erzeugten Abbildungen, in denen die Grundfrequenzverläufe grafisch dargestellt werden, ist die Frequenz (Pitch) in Hertz (Hz) auf der y-Achse von 0 – bis 600 Hz dargestellt, der Zeitverlauf (Time) in Sekunden liegt auf der x-Achse. Dabei ist zu beachten, dass die Abbildungen immer dieselbe Breite haben, ungeachtet ob ein Zeitverlauf von sechs oder 26 Sekunden dargestellt wird. Eine vergleichende Interpretation nach optischen Eindrücken ist daher nur bedingt möglich. Der eher abgehackt erscheinende Verlauf der Grundfrequenz entsteht durch die genaue Messung mit einem Analysefenster (frame) von 0,01 Sekunden, wodurch, vor allem artikulationsbedingt, auch tonlose Bereiche (voiceless) sichtbar werden. Der besseren Lesbarkeit halber sind die Messpunkte mit Linien verbunden.

Bei der Sprechgeschwindigkeit sind die Berechnungen mit zwei Nachkommastellen sinnvoll, da die Ergebnisse teilweise eng beisammen liegen und so eine bessere Differenzierung möglich ist. Bei der Analyse der Grundfrequenzen gibt Praat bei Range und Interquantilsabstand die Rechenergebnisse mit einer Nachkommastelle an. Um einheitliche Aussagen treffen zu können, werden daher nun alle Werte, die hier in Hertz und in Halbtönen angegeben sind, auf eine Nachkommastelle gerundet. Dies ist bei den Distanzen der Frequenzen angemessen aussagekräftig.

Diese Abstände werden nicht nur in Hertz sondern auch in Halbtönen angegeben, da diese, aus der Musik kommende und auditiv wahrnehmungsbezogene Intervallbezeichnung unabhängig von Frequenzlagen, beziehungsweise Tonhöhen, ist. Sie ist daher gut geeignet, Range und Interquantilsabstand bei verschiedenen Stimmen zu vergleichen. Im folgenden wird darauf noch genauer eingegangen werden.

Der Spot **Möbelix** (Abb. 7) kann nicht in der Kategorie Mann- oder Frauenstimme betrachtet werden, da hier, wie schon angesprochen, eine Männerstimme tontechnisch verändert wurde. Der Stimmklang entspricht einer typischen Comicfigur. In den Tabellen wird daher die Stimme mit Comic bezeichnet. Dies beeinflusst jedoch nicht die Fragestellung hinsichtlich der Aufmerksamkeits- bzw. Kompetenzsignale. Der berechnete Minimumwert der Grundfrequenz beträgt 126,3 Hz, der Maximumwert 337,7 Hz, die Differenz dieser zwei Werte (Range) 211,4 Hz. Der Mittelwert (Average) liegt auf der Höhe von 248,1 Hz. Die Standardabweichung liegt im unteren Bereich von 63,2 Hz oder 4,9 Halbtönen. Der Interquantilsabstand zwischen dem 10% und 90% Quantil, bei dem die oberen und die unteren zehn Prozent der Werte nicht berücksichtigt werden, beträgt 169,8 Hz. Dies bedeutet ein Intervall von 13,1 Halbtönen.

Der von einer Männerstimme gesprochene Spot **Alpecin** (Abb. 8) weist zwei kleine Ausreißer in der zweiten und in der siebten Sekunde auf, wodurch der Minimumwert der Grundfrequenz bei 34,3 Hz und der Maximumwert bei 556,2 Hz liegt, was einen sehr hohen Range von 521,9 Hz

ergibt. Der Interquantilsabstand (10-90%) beträgt jedoch nur 88,6 Hz oder 11,6 Halbtöne und kommt so dem tatsächlichen Grundfrequenzverlauf wesentlich näher. Die durchschnittliche Höhe

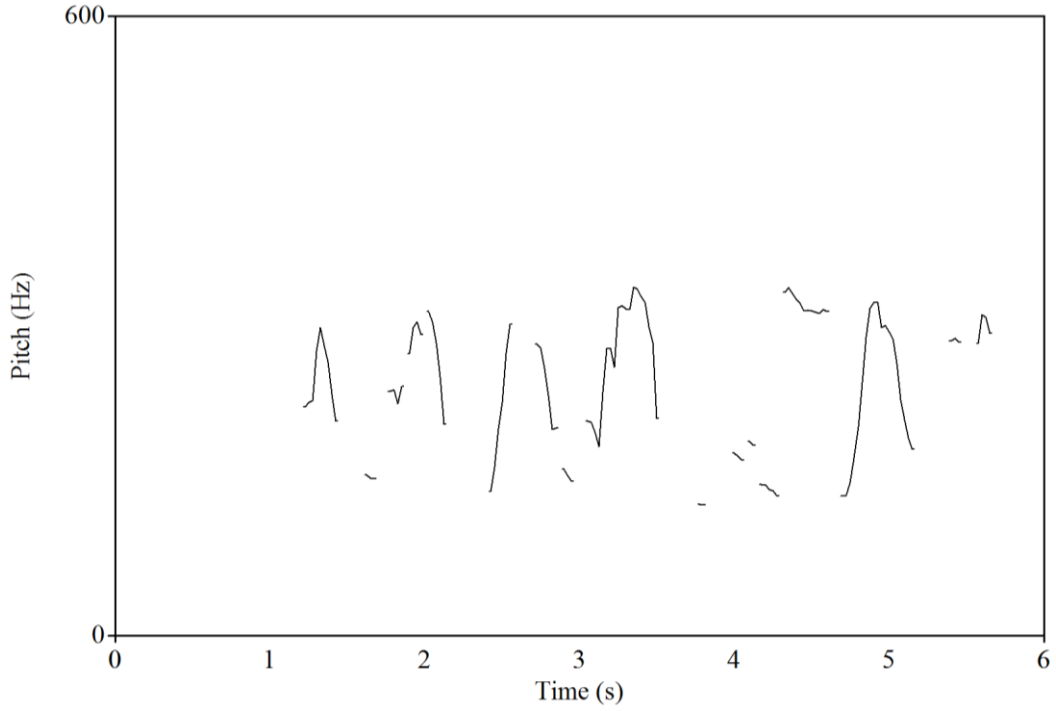


Abbildung 7: Spot Möbelix: Grundfrequenzverlauf

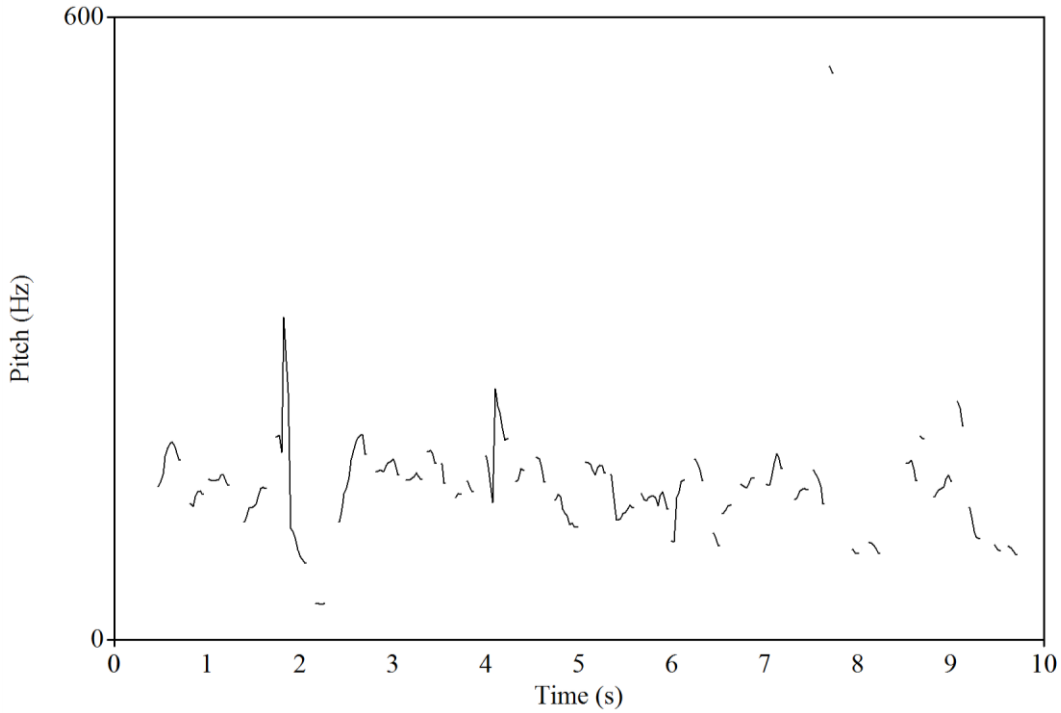


Abbildung 8: Spot Alpecin: Grundfrequenzverlauf

liegt bei 146,6 Hz, die Standardabweichung bei 51,7 Hz oder 5,6 Halbtönen. Das akustische Signal in der ersten Sekunde, das gleichzeitig mit der Stimme zu hören ist, ist in der grafischen Darstellung nicht sichtbar. Da die bestimmenden Kennzahlen zur Grundfrequenzmessung dadurch nicht verändert werden, ist dieses Signal vernachlässigbar. Die markante vertikale Linie zwischen 1,6 und 2,3 Sekunden, die von rund 310 Hz auf etwa 61 Hz abfällt, entsteht bei dem betonten Wort „tunen“. Der Plosivlaut „t“ und der anschließende betonte, und wie das „i“ geschlossene, höherfrequente Vokal „u“ verursachen das abrupte Ansteigen der Signalkurve. Durch die Endsilbe „en“, die unbetont ist und abphrasiert wird, entsteht der markante Abfall.

Der Spot **Neuroth** (Abb. 9), der ebenfalls von einer Männerstimme gesprochen wird, hat eine ähnliche Charakteristik. Durch einen Ausreißer bei dem zwei Mal hintereinander gesprochenem „t“ bei der Textstelle „Hörgerät trägt“ in der sechsten und siebten Sekunde ist der Range zwischen niedrigstem (47,8 Hz) und höchstem Wert (586,4 Hz) 538,6 Hz. Der Interquantilsabstand (10-90%) beträgt jedoch 100,7 Hz oder 17,7 Halbtöne und kommt dem tatsächlichen Grundfrequenzverlauf wieder näher. Der Mittelwert liegt bei 122,7 Hz, die Standardabweichung bei 94,0 Hz oder 8,9 Halbtönen.

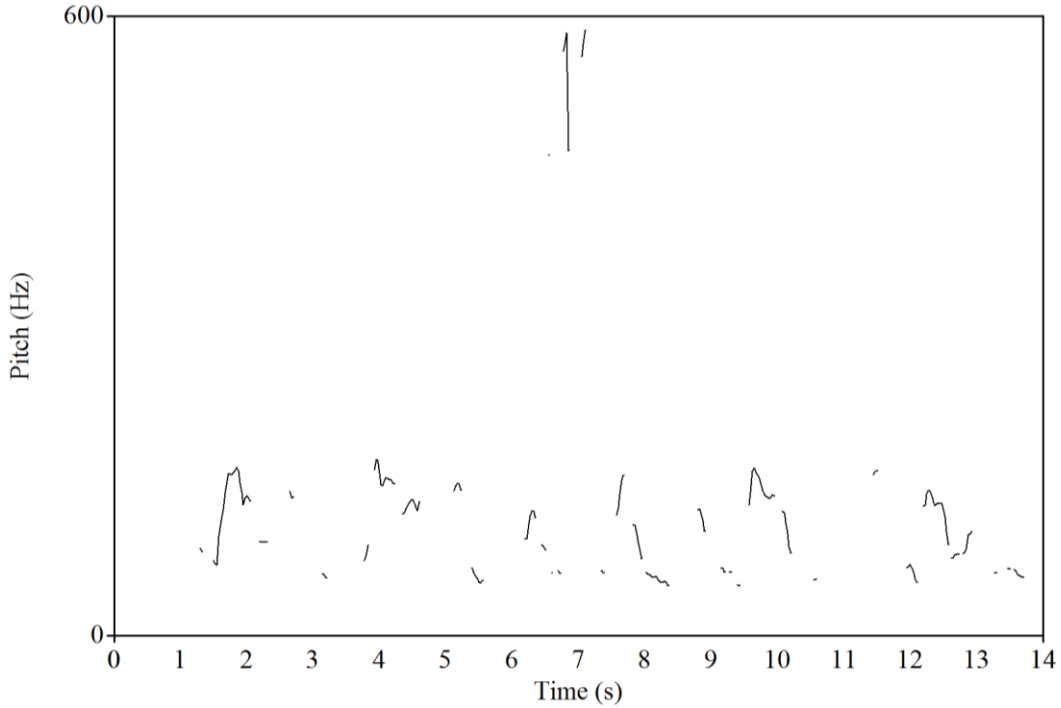


Abbildung 9: Spot Neuroth: Grundfrequenzverlauf

Der Spot **WC-Ente** (Abb. 10) wird auch von einer Männerstimme gesprochen. Der Mittelwert der Grundfrequenz ist bei 155,1 Hz. Die Signale in der oberen Hälfte der Grafik liegen über 400 Hz und entstehen vor allem durch den Zischlaut „s“. Sie liegen damit weit über dem 90% Quantil von 218,7 Hz. Der Interquantilsabstand (10-90%) beträgt 124,1 Hz oder 14,5 Halbtöne, der Range

hingegen reicht von 66,6 Hz bis 590,7 Hz und hat den Wert von 524,1 Hz. Die Standard-abweichung beträgt 106,1 Hz oder 7,9 Halbtöne.

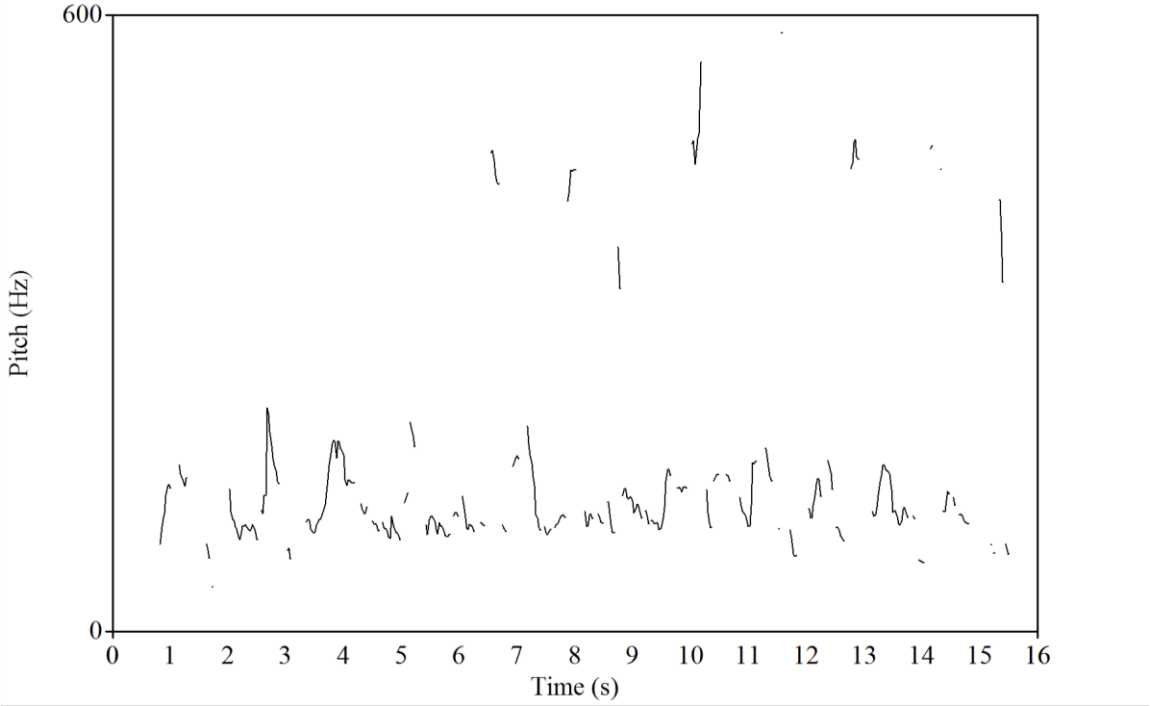


Abbildung 10: Spot WC-Ente: Grundfrequenzverlauf

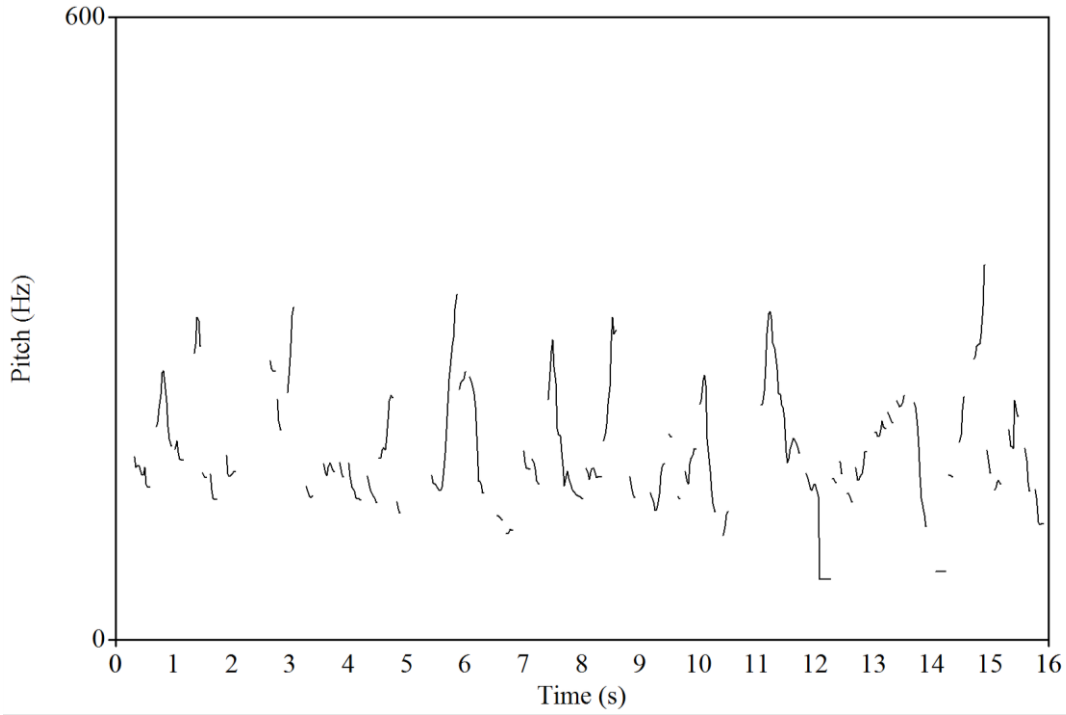


Abbildung 11: Spot Canesten Glutrimazol: Grundfrequenzverlauf

Der Spot **Canesten Glutrimazol** (Abb. 11) wird von einer Frauenstimme gesprochen, was auch mit freiem Auge an der insgesamt höheren Lage des Grundfrequenzverlaufs erkennbar ist. Der Maximumwert der Grundfrequenz liegt hier bei 332,6 Hz, der Minimumwert bei 58,3 Hz. Der kleinere Ausreißer nach unten im Bereich der 12. Sekunde bei der Textstelle „wohlfühlen mit“ entsteht durch die Abfolge der zwei Konsonanten „n“ und „m“. Der Range beträgt 274,3 Hz, der Interquantilsabstand (10-90%) 134,2 Hz oder 12,6 Halbtöne. Der Mittelwert der Grundfrequenz liegt bei 180,5 Hz, die Standardabweichung bei 55,5 Hz oder 5,7 Halbtönen. Die leise Hintergrundmusik zwischen der 11. und 15. Sekunde ist auch hier im Grundfrequenzverlauf nicht erkennbar und daher für die vorliegenden Analysen vernachlässigbar.

Der Spot **Plantur** (Abb. 12) wird ebenfalls von einer Frauenstimme gesprochen. Der Minimumwert der Grundfrequenz beträgt hier 97,9 Hz, der Maximumwert 323,3. Es gibt keine markanten Ausreißer, der Range ist daher mit 225,4 Hz im Vergleich niedrig. Der Interquantilsabstand (10-90%) liegt bei 103,9 Hz oder 10,6 Halbtönen. Der am höchsten, nämlich knapp um 300 Hz, liegende Bereich in der 17. und 18. Sekunde wird durch den Vokal „i“ in den Worten „die Koffein (-therapie)“ erzeugt. Der Mittelwert der Grundfrequenz liegt bei 163,3 Hz, die Standardabweichung ist mit 40,9 Hz oder 4,0 Halbtönen im Vergleich zu den anderen Spots sehr niedrig. Gut sichtbar sind hier insbesondere die ersten drei Pausen in der fünften, neunten und 15. Sekunde.

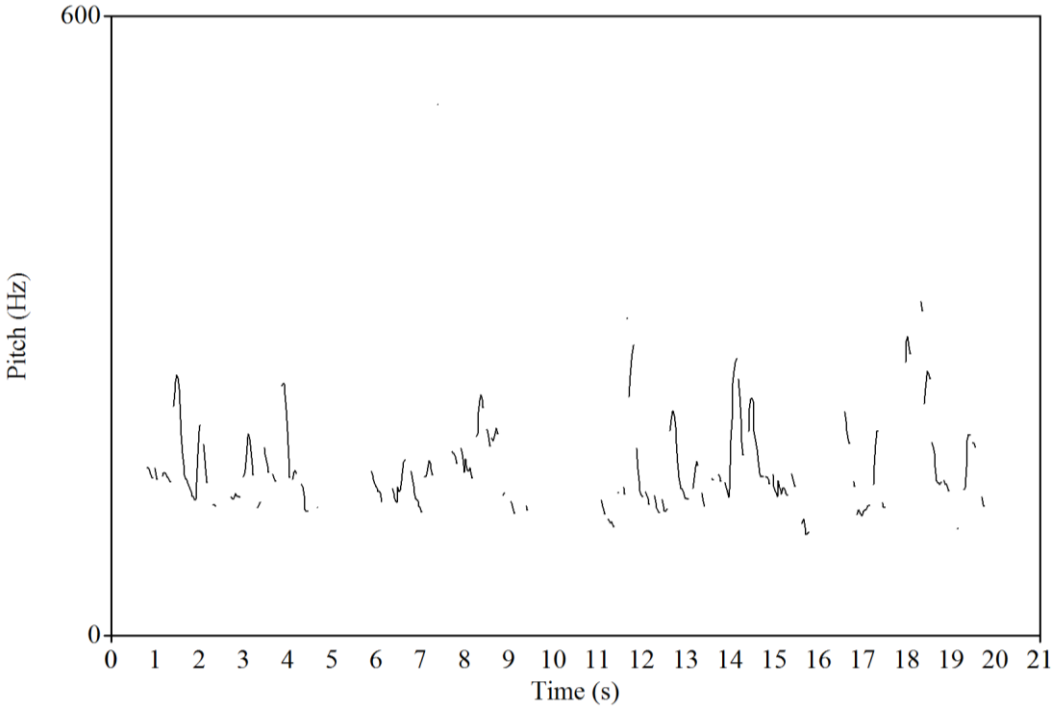


Abbildung 12: Spot Plantur: Grundfrequenzverlauf

Der Spot **Sensodyne** (Abb. 13), der auch von einer Frauenstimme gesprochen wird, liegt insgesamt höher als die anderen. Dies ist auch am Wert des 10% Quantils erkennbar, der mit 172,1 Hz deutlich höher als bei den anderen Spots liegt. Der markante Abfall in der 16. Sekunde entsteht bei dem Wort „tun“ durch Abphrasierung vor der Pause. Durch die Ausreißer in der siebenten und 16. Sekunde liegt der Minimumwert bei 83,6 Hz, der Maximumwert bei 504,3 Hz. Dieser hohe Wert entsteht insbesondere durch eine starke Betonung des ersten Vokals „i“ im Wort „das Wichtigste“ in der achten Sekunde. Der Range erreicht dadurch den Wert von 420,7 Hz, der Interquantilsabstand (10-90%) beträgt jedoch nur 111,7 Hz oder 8,7 Halbtöne. Der Mittelwert der Grundfrequenz liegt bei 221,4 Hz, die Standardabweichung bei 54,8 Hz oder 4,1 Halbtönen.

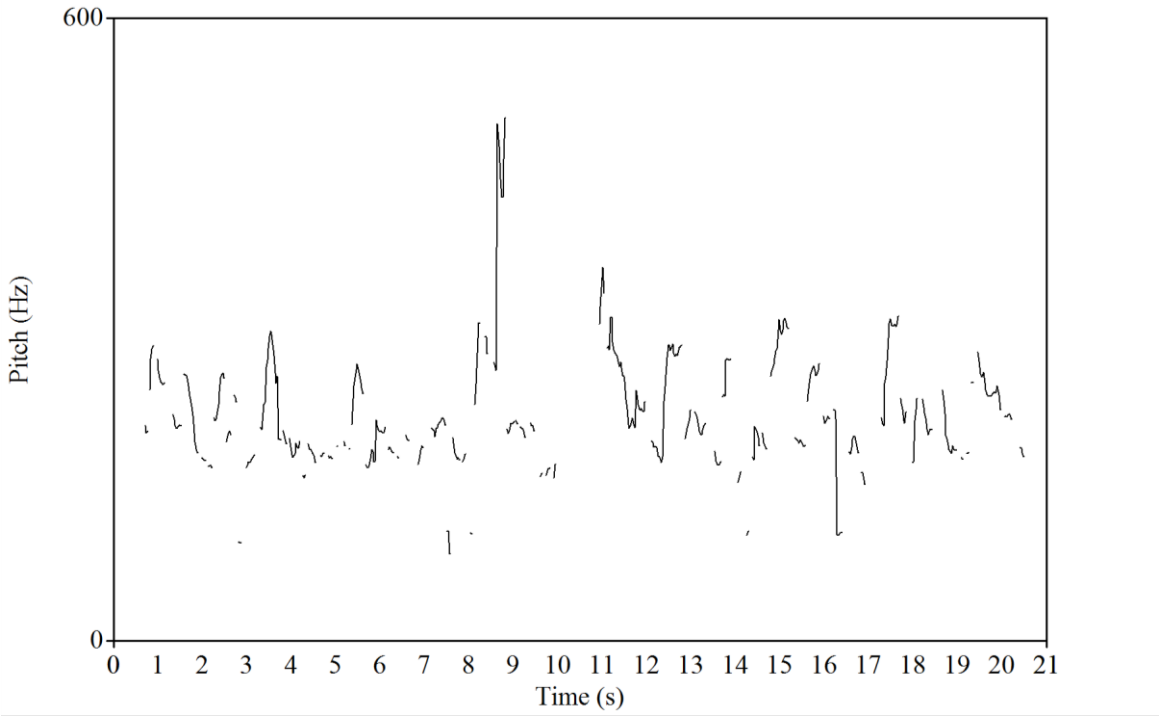


Abbildung 13: Spot Sensodyne: Grundfrequenzverlauf

Der Spot **Canesten Bifonazol** (Abb. 14) wird von einer tiefen Männerstimme gesprochen. Der Minimumwert der Grundfrequenz liegt bei 32,2 Hz, das 10% Quantil bei 58,7 Hz. Durch einige Ausreißer im obersten Frequenzbereich, die durch die Zischlaute „s“ und „z“ wie zum Beispiel beim Wort „Fußpilz“ erzeugt werden, liegt der Maximumwert bei sehr hohen 596,5 Hz. Der Range beträgt dadurch 564,3 Hz, der Interquantilsabstand (10-90%) hingegen nur im Vergleich niedrige 71,1 Hz oder 13,8 Halbtöne. Der durchschnittliche Wert der Grundfrequenz liegt bei 112,6 Hz, die Standardabweichung bei 107,9 Hz oder 8,8 Halbtönen.

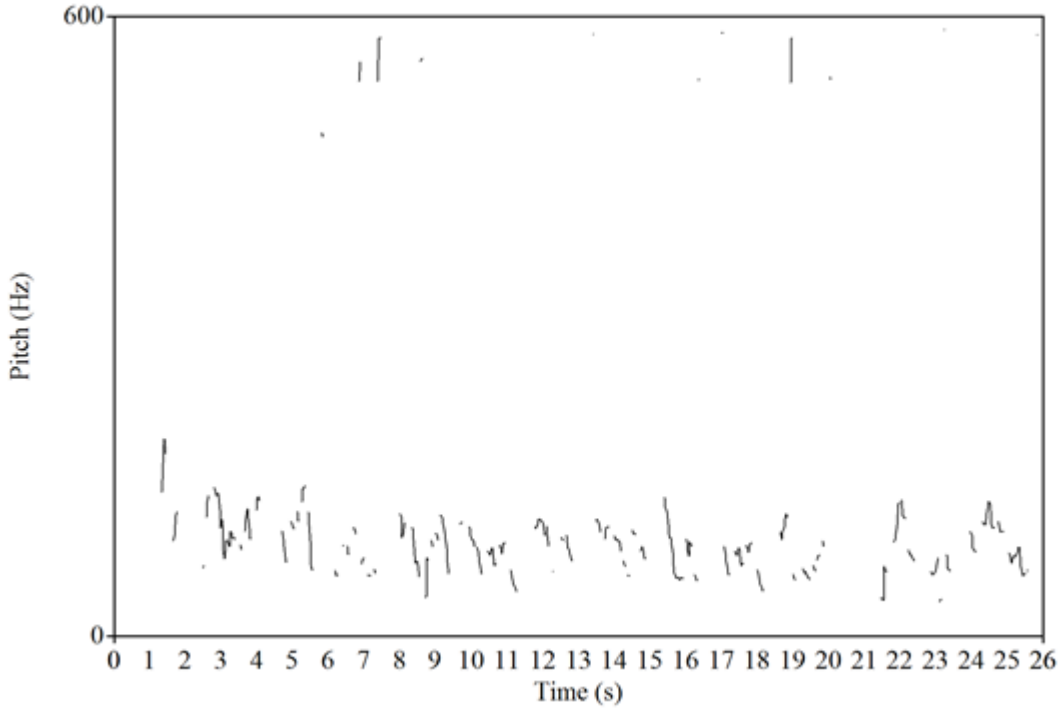


Abbildung 14: Spot Canesten Bifonazol: Grundfrequenzverlauf

In Tabelle 4 werden nun die Werte zu den Eigenschaften der Grundfrequenzen der Spots in einer Übersicht zusammengefasst. Die Reihung erfolgt hier aufsteigend nach dem Mittelwert. Neben den Minimum- und Maximumwerten sind auch Range, Interquantilsabstand (10-90%), Standardabweichung sowie das Geschlecht der Stimme dargestellt. Die Frauenstimmen sowie die hoch liegende Stimme von Möbelix werden zur leichteren optischen Unterscheidung grau hinterlegt.

Werbespot	**Min.**	**Max.**	**Range**	**10-90%**	**Mittelwert**	**Standard-abweichung**	**Stimme**
Can. Bi.	32,2	596,5	564,3	71,1	**112,6**	107,9	Mann
Neuroth	47,8	586,4	538,6	100,7	**122,7**	94,0	Mann
Alpecin	34,3	556,2	521,9	88,6	**146,6**	51,7	Mann
WC-Ente	66,6	590,7	524,1	124,1	**155,1**	106,1	Mann
Plantur	97,9	323,3	225,4	103,9	**163,3**	40,9	Frau
Can. Gl.	58,3	332,6	274,3	134,2	**180,5**	55,5	Frau
Sensodyne	83,6	504,3	420,7	111,7	**221,4**	54,8	Frau
Möbelix	126,3	337,7	211,4	169,8	**248,1**	63,2	Comic

Tabelle 4: Übersicht über die Eigenschaften der Grundfrequenzen (alle Werte in Hz)

Bei der Reihung nach dem Mittelwert wird auch der bereits in Kapitel 2.1 angesprochene Unterschied der tiefer liegenden Männerstimmen zu den höheren Frauenstimmen deutlich. Wobei an dieser Stelle vorausgreifend angemerkt werden soll, dass die geschlechtsspezifische Differenz auffallend gering ist, da insbesondere die Frauenstimmen der Spots Plantur und Canesten Glutimazol bemerkenswert tief liegen.

Der Mittelwert, der den Durchschnitt aller gemessenen Werte darstellt, ist natürlich auch von den, durch die Artikulation bedingten, Ausreißern beeinflusst. Dieser Einfluss wird mittels der Berechnung des Medians (50% Quantil) begrenzt, der so platziert wird, dass jeweils 50% der Daten unter- beziehungsweise oberhalb dieses Lagemaßes liegen. Dadurch können die Stimmqualitäten präziser beschrieben werden. In Tabelle 5 werden Mittelwert und Median aufsteigend gelistet. Die Rangreihen der Spots sind fast ident, es gibt nur einen Tausch bei Rang drei und vier (Alpecin und WC-Ente). Die Werte des Medians liegen jedoch bei sechs Spots deutlich unter denen des Mittelwerts. Nur beim Spot Alpecin sind die Werte fast ident, beim Spot Möbelix ist der Median sogar deutlich höher als der Mittelwert. Nachdem jedoch der Mittelwert in der Literatur oft zur Beschreibung der Grundfrequenz angegeben wird, wird dieses Lagemaß zur besseren Vergleichbarkeit mit anderen Forschungsergebnissen auch weiterhin berücksichtigt.

Werbespot	**Mittelwert (Hz)**	**Rang**	**Median (Hz)**	**Rang**	**Stimme**
Canesten Bi.	**112,6**	1	89,8	1	Mann
Neuroth	**122,7**	2	112,2	2	Mann
Alpecin	**146,6**	3	147,1	4	Mann
WC-Ente	**155,1**	4	115,6	3	Mann
Plantur	**163,3**	5	152,9	5	Frau
Canesten Gl.	**180,5**	6	165,7	6	Frau
Sensodyne	**221,4**	7	209,6	7	Frau
Möbelix	**248,1**	8	266,5	8	Comic

Tabelle 5: Mittelwert und Median der Grundfrequenzen im Vergleich

In den grafischen Darstellungen zu den Grundfrequenzverläufen wird deutlich, dass die Minimum- und Maximumwerte nur bedingte Aussagekraft hinsichtlich der Eigenschaften der Grundfrequenz besitzen. Es erscheint daher sinnvoll statt des Range, der aus der Differenz dieser zwei Werte errechnet wird, den Interquantilsabstand (10-90%) zur Bewertung zu verwenden, da hier die Ausreißer unberücksichtigt bleiben. Es soll an dieser Stelle noch einmal betont werden, dass

aufgrund der Komplexität der vorliegenden Messdaten die Berechnung diverser Parameter nach speziellen Algorithmen erfolgen muss. Im Analyseprogramm Praat wird daher auch von „Estimated Quantile“ beziehungsweise von „Estimated Spreading“ zwischen dem 10% und dem 90% Quantil gesprochen.[61] In Tabelle 6 werden die Werbespots aufsteigend nach diesem Interquantilsabstand (Estimated Spreading, 10-90%) gereiht. Zum vertieften Verständnis werden auch die in Praat standardmäßig errechneten 16% und 84% Quantile sowie der Median (50% Quantil) in Tabelle 6 dargestellt.

Werbespot	**Quantil (Hz)**					**Interquantilsabstand (Hz)**	**Rang**
	10%	**16%**	**50%**	**84%**	**90%**	**10 – 90 %**	
Can. Bi.	58,7	62,4	89,8	116,5	129,8	**71,1**	1
Alpecin	92,9	108,6	147,1	173,1	181,5	**88,6**	2
Neuroth	56,9	59,5	112,2	151,5	157,4	**100,7**	3
Plantur	122,7	127,5	152,9	198,6	226,5	**103,9**	4
Sensodyne	172,1	177,3	209,6	271,2	283,8	**111,7**	5
WC-Ente	94,7	99,3	115,6	170,4	218,7	**124,1**	6
Can. Gl.	125,5	137,4	165,7	240,9	259,7	**134,2**	7
Möbelix	149,9	169,8	266,5	315,2	319,4	**169,8**	8

Tabelle 6: Quantile und Interquantilsabstand 10-90% in Hz (Frauenstimmen grau hinterlegt)

In den vorhergehenden Beschreibungen der Spots werden Interquantilsabstand und Standardabweichung in Hertz und in Halbtönen[62] angegeben. Dies ist hier insofern relevant, als die Größe eines Intervalls gemessen in Hz abhängig von der Lage der Frequenzzahlen ist, in dem es sich befindet.[63] Liegt zum Beispiel die Standardabweichung in einem höheren Frequenzbereich, ergibt der jeweilige Wert in Hertz ein geringeres Intervall als in einem tieferen Frequenzbereich. Die Ergebnisse in Hz haben daher bei tiefen Stimmen eine andere Aussagekraft als bei hohen. Mit dem

[61] Praat liefert bei den Interquantilsabständen in den Outputlisten gerundete Ergebnisse, auf die in den vorliegenden Tabellen auch Bezug genommen wird. Beim Vergleich von händischer Berechnung und interner Berechnung auf Maschinengenauigkeit können dadurch geringfügige Diskrepanzen in den Nachkommastellen auftreten.

[62] Zu Halbtönen siehe auch Fußnote 55 in Kapitel 3.3.

[63] Wenn „a^1“ auf 440 Hz gestimmt ist und man 12 Halbtonschritte (oder eine Oktave) hinauf geht, liegt „a^2“ auf 880 Hz. Diese 12 Halbtonschritte hinauf erstrecken sich daher über den Frequenzbereich von 440 Hz. Geht man hingegen 12 Halbtonschritte (oder eine Oktave) hinunter liegt „a“ auf 220 Hz und die 12 Halbtonschritte erstrecken sich über einen Frequenzbereich von 220 Hz. Dementsprechend hat auch jeder einzelne Halbtonschritt einen spezifischen Frequenzbereich, der mit der jeweiligen Höhe korreliert.

aus der Musik kommenden Halbton-Intervall, das gegen die Lage der Frequenzen resistent ist, kann dies jedoch relativiert werden und die Ergebnisse werden dadurch vergleichbar. Zur Illustration werden in Tabelle 7 die Ergebnisse für den Interquantilsabstand (10-90%) in Hertz und im Halbton-Intervall einander gegenübergestellt. Um den Einfluss der jeweiligen Frequenzlagen zu verdeutlichen, erfolgt die erste Reihung nach dem Mittelwert.

Werbespot	**Mittelwert (Hz)**	**Rang**	**Interquantils-abstand 10 – 90% (Hz)**	**Rang**	**Interquantils-abstand 10 – 90% (Halbtöne)**	**Rang**
Can. Bi.	112,6	1	**71,1**	1	**13,8**	6
Neuroth	122,7	2	**100,7**	3	**17,7**	8
Alpecin	146,6	3	**88,6**	2	**11,6**	3
WC-Ente	155,1	4	**124,1**	6	**14,5**	7
Plantur	163,3	5	**103,9**	4	**10,6**	2
Can. Gl.	180,5	6	**134,2**	7	**12,6**	4
Sensodyne	221,4	7	**111,7**	5	**8,7**	1
Möbelix	248,1	8	**169,8**	8	**13,1**	5

Tabelle 7: Mittelwert, Interquantilsabstand 10-90% in Hz und in Halbtönen (Frauenstimmen grau hinterlegt)

Die Abhängigkeit der Intervallgröße von der Frequenzlage beziehungsweise der Stimmhöhe wird zum Beispiel bei Neuroth, Plantur und Sensodyne gut deutlich. Gemessen in Hertz ist der Interquantilsabstand (10-90%) bei allen drei Spots in einem ähnlichen Bereich (100,7 Hz, 103,9 Hz, 111,7 Hz). Dies ergibt bei der tiefen Männerstimme von Neuroth ein hohes Intervall von 17,7 Halbtönen, bei der tiefen Frauenstimme von Plantur 10,6 Halbtöne und bei der höchsten dieser drei Stimmen, Sensodyne, nur ein Intervall von 8,7 Halbtönen. Auch beim Vergleich des tiefsten Spots Canesten Bifonazol mit dem höchsten Spot Möbelix wird diese Abhängigkeit deutlich. Beide haben bei Mittelwert und Interquantilsabstand (10-90%) in Hz den jeweils tiefsten beziehungsweise höchsten Wert. Doch das Halbtonintervall ist mit etwas über 13 Halbtönen fast ident.

Der Mittelwert als Maßzahl für die Stimmhöhe und der Range beziehungsweise Interquantilsabstand sowie die Standardabweichung als Maßzahlen für den Ausdruck und die Lebendigkeit oder die Monotonie einer Stimme stehen in engem Zusammenhang. Um sinnvolle Aussagen zu den Stimmqualitäten treffen zu können, muss auch berücksichtigt werden, ob eine Frauenstimme hoch oder tief ist und die anderen Maßzahlen dazu in Bezug gesetzt werden. Dasselbe gilt für Männer-

stimmen. Eine Rangreihung aller acht Spots hat daher hier keine Aussagekraft mehr.[64] In Tabelle 8 werden daher die Männer- und Frauenstimmen getrennt. Der Spot Möbelix wird dabei aufgrund des hohen Mittelwertes den Frauenstimmen zugeteilt. Hier ist nun deutlich zu sehen, dass die Rangreihungen innerhalb der Stimmgruppen wesentlich homogener sind, als man es nach Tabelle 7 interpretieren würde. Die Reihungen unterscheiden sich in beiden Gruppen nur jeweils um maximal einen Rang.

Werbespot	**Mittelwert (Hz)**	**Rang**	**Interquantils-abstand 10 – 90% (Hz)**	**Rang**	**Interquantils-abstand 10 – 90% (Halbtöne)**	**Rang**
Can. Bi.	112,6	1	**71,1**	1	**13,8**	2
Neuroth	122,7	2	**100,7**	3	**17,7**	4
Alpecin	146,6	3	**88,6**	2	**11,6**	1
WC-Ente	155,1	4	**124,1**	4	**14,5**	3
Plantur	163,3	1	**103,9**	1	**10,6**	2
Can. Gl.	180,5	2	**134,2**	3	**12,6**	3
Sensodyne	221,4	3	**111,7**	2	**8,7**	1
Möbelix	248,1	4	**169,8**	4	**13,1**	4

Tabelle 8: Mittelwert, Interquantilsabstand 10-90% in Hz und in Halbtönen, Rangreihen getrennt nach Frauen- und Männerstimmen (Frauenstimmen grau hinterlegt)

Die dritte Maßzahl zur Beschreibung der Grundfrequenz ist die Standardabweichung, mit der die Streuung der Werte um ihr Mittel beschrieben wird. In Kapitel 3.2.2. wurde bereits darauf hingewiesen, dass diese Ergebnisse von den Ausreißern beeinflusst werden. Dies muss daher bei der Analyse und Interpretation derselben berücksichtigt werden.

In Tabelle 9 werden die Spots aufsteigend nach den Mittelwerten gereiht und die Standard-abweichung in Hertz und in Halbtönen mit den jeweiligen Rangreihen dargestellt. Dabei wird insbesondere bei den Spots Canesten Bifonazol und Neuroth auch wieder die Abhängigkeit der in Hz gemessenen Intervallgröße, in dem Fall der Streuung um den Mittelwert, von der Stimmhöhe deutlich. Der Mittelwert dieser zwei Spots unterscheidet sich deutlich um fast genau 10 Hz. Doch während bei der Standardabweichung in Hz eine Differenz von fast 14 Hz zu sehen ist, unter-scheidet sich das Halbton-Intervall hingegen nur um eine Nachkommastelle.

[64] Vergleiche dazu auch: Enterlein/Bartels/Sendlmeier (2005:23).

Werbespot	Mittelwert (Hz)	Rang	Standard-abweichung (Hz)	Rang	Standard-abweichung (Halbtöne)	Rang
Canesten Bi.	112,6	1	**107,9**	4	**8,8**	3
Neuroth	122,7	2	**94,0**	2	**8,9**	4
Alpecin	146,6	3	**51,7**	1	**5,6**	1
WC-Ente	155,1	4	**106,1**	3	**7,9**	2
Plantur	163,3	1	**40,9**	1	**4,0**	1
Canesten Gl.	180,5	2	**55,5**	3	**5,7**	4
Sensodyne	221,4	3	**54,8**	2	**4,1**	2
Möbelix	248,1	4	**63,2**	4	**4,9**	3

Tabelle 9: Mittelwert, Standardabweichung der Grundfrequenzen in Hz und in Halbtönen, Rangreihen getrennt nach Frauen- und Männerstimmen (Frauenstimmen grau hinterlegt)

Im folgenden Kapitel werden die Ergebnisse dieser Berechnungen zusammengefasst und im Kontext relevanter Ergebnisse der Forschungsliteratur diskutiert. Im fünften Abschnitt werden sie hinsichtlich der adressierten Forschungsfragen dieser Arbeit im Detail besprochen.

4.4. ZUSAMMENFASSUNG UND SIGNALANALYSE

Die Sprechgeschwindigkeit wurde bei insgesamt 13 Spots berechnet (siehe Tabelle 2 und 3). Die globale Sprechgeschwindigkeit dieser Spots liegt zwischen 3,38 (Leiner) und 5,63 (Möbelix) Silben pro Sekunde.

In der Literatur konnten keine Referenzwerte zur Sprechgeschwindigkeit in Werbespots gefunden werden, jedoch gibt es diesbezügliche Studien im Zusammenhang mit Nachrichtensprechern. In einer perzeptiven und akustischen Analyse deutscher Nachrichtensprecher messen Schubert/Sendlmeier (2005) Sprechgeschwindigkeiten zwischen 5,1 und 6,47 Silben pro Sekunde unter Berücksichtigung der Pausenzeiten. Straßner (1982, in: Schubert/Sendlmeier 2005:18) definiert dafür ein wesentlich langsameres Maximalmaß von 4,17 Silben pro Sekunde. Fiukowski (1999, in: Schubert/Sendlmeier 2005:18) legt als Grenze für zu schnelles Sprechen wiederum 5,8 Silben pro Sekunde fest. Dieser Wert entspricht auch den Ergebnissen der perzeptiven Analyse von Schubert/Sendlmeier (2005), in denen Sprachproben mit einer Geschwindigkeit ab 5,88 Silben pro Sekunde als schnell beziehungsweise als zu schnell bewertet wurden.

Da bei allen genannten Werten die Pausenzeiten nicht explizit definiert sind, ist kein direkter Vergleich mit den vorliegenden Ergebnissen aus den Werbespots möglich. Trotzdem ist eindeutig feststellbar, dass hier wesentlich langsamer gesprochen wird. Nur fünf Spots (VW, Plantur, Alpecin, Sensodyne, Möbelix) erreichen unter Berücksichtigung der Pausen Werte von knapp über 5,1 Silben pro Sekunde. Auch der schnellste Spot Möbelix liegt mit 5,63 Silben noch unter der oben definierten Grenze von 5,8 Silben.

Im Kontext der Alltagskommunikation wird eine Sprechgeschwindigkeit von etwa sieben Silben pro Sekunde als schnell bewertet (Enterlein/Bartels/Sendlmeier 2005:33). Mit einer globalen Sprechgeschwindigkeit von 3,38 bis 3,89 Silben pro Sekunde liegen vier Spots (Leiner, Cafe Haag, Neuroth, Innocent) rund 50% unter diesem Wert und können somit durchaus als langsam bezeichnet werden. Wobei daran erinnert werden soll, dass diese Spots, abgesehen von Neuroth, zusätzliche akustische Signale und Hintergrundgeräusche beinhalten. Dadurch kann die langsame Sprache etwas ausgeglichen werden und in Summe ein anderer Höreindruck entstehen.

Zusammenfassend kann auf Basis dieser Daten festgestellt werden, dass in keinem Werbespot die Grenzen für schnelles beziehungsweise zu schnelles Sprechen über 5,8 Silben pro Sekunde erreicht werden. Eine globale Sprechgeschwindigkeit von unter 4 Silben pro Sekunde kann hingegen in Relation zu den oben genannten Werten aus der Alltagskommunikation sowie auch dem definierten Grenzwert von Straßner (1982) als langsam bezeichnet werden. Spots mit Werten von 4,1 bis 4,9 Silben pro Sekunde werden darauf bezugnehmend im folgenden als Spots mit mittlerer Sprechgeschwindigkeit bezeichnet. Diejenigen mit über 5 Silben pro Sekunde als eher schnell.

Für die Berechnung der Eigenschaften der Grundfrequenz wurden die Spots Leiner, Cafe Haag, Innocent, Nestea und VW ausgeschieden, da die Frequenzen der Hintergrundgeräusche nicht von denjenigen der Sprechstimme getrennt werden konnten. Die Ergebnisse, die für die restlichen acht Spots errechnet wurden, sind in Tabelle 10 dargestellt. Darin werden der Mittelwert in Hz, Interquantilsabstand (10-90%) und Standardabweichung in Halbtönen sowie die globale Sprechgeschwindigkeit der Spots mit den jeweiligen Rängen zusammengefasst. Die Reihung nach dem Mittelwert entspricht einer Ordnung von der tiefsten zur höchsten Stimme, wobei die ersten vier Spots von Männerstimmen gesprochen werden, die nächsten drei von Frauenstimmen und die letzte und höchste von einer Comic-Stimme.

Wie in Kapitel 2.1. ausführlicher dargestellt, sprechen Männer im allgemeinen in Frequenzbereichen von 100 bis 180 Hz, Frauen von 150 bis 220 Hz (Mayer 2010). Terhardt (1998) und Goldstein (2002) geben den Mittelwert der männlichen Sprechstimme mit 120 Hz an, den Mittelwert der weiblichen Sprechstimme bei 240 beziehungsweise 210 Hz. Die Stimme des Spots Canesten Bifonazol liegt mit 112,6 Hz im untersten Bereich, Neuroth ist nah bei 120 Hz, Alpecin und insbesondere WC-Ente liegen mit 146,6 Hz und 155,1 Hz schon deutlich im höheren Bereich der Männerstimmen. Die Frauenstimmen liegen hingegen eher unter den genannten Mittelwerten.

Bei den Spots von Plantur und Canesten Glutrimazol liegen die Stimmen in der unteren Hälfte des von Mayer (2010) genannten Bereichs, Sensodyne bewegt sich mit 221,4 Hz im oberen Bereich. Nur die Comic-Stimme von Möbelix liegt etwas über dem von Terhardt (1998) genannten Mittelwert von 240 Hz. Zusammenfassend kann man sagen, dass bei zwei Spots die Männerstimmen hoch, und bei zwei Spots die Frauenstimmen tief sind. Nur eine Männerstimme kann eindeutig als tief, die Comic-Stimme, die den Frauenstimmen zugeordnet ist, kann wiederum eindeutig als hoch bezeichnet werden.

Werbespot	**Mittelwert (Hz)**	**Rang**	**Interquantils-abstand 10 – 90% (Halbtöne)**	**Rang**	**Standard-abweichung (Halbtöne)**	**Rang**	**Silben/Sek. global**	**Rang**
Can. Bi.	112,6	1	13,8	2	8,8	3	4,03	2
Neuroth	122,7	2	17,7	4	8,9	4	3,39	1
Alpecin	146,6	3	11,6	1	5,6	1	5,21	4
WC-Ente	155,1	4	14,5	3	7,9	2	4,48	3
Plantur	163,3	1	10,6	2	4,0	1	4,79	2
Can. Gl.	180,5	2	12,6	3	5,7	4	4,78	1
Sensodyne	221,4	3	8,7	1	4,1	2	5,34	3
Möbelix	248,1	4	13,1	4	4,9	3	5,63	4

Tabelle 10: Überblick über die Eigenschaften der Grundfrequenz sowie die Sprechgeschwindigkeit mit den Rangreihen getrennt nach Frauen- und Männerstimmen (Frauenstimmen grau hinterlegt)

Statt des Range, das heißt, der Differenz des höchsten und des niedrigsten Wertes der Grundfrequenz wird in der vorliegenden Arbeit der Abstand des 10% und des 90% Quantils verwendet, da hier artikulationsbedingte Ausreißer eliminiert werden können. Zur Berechnung des Range bei Vorliegen von Ausreißern werden in der Literatur unterschiedliche oder keine genauen Angaben gemacht. Vergleiche mit anderen Forschungsergebnissen sind daher nur unter Berücksichtigung einer möglichen Unschärfe möglich.

Nach Enterlein/Bartels/Sendlmeier (2005:25), die in ihrer Studie den Interquantilsabstand zwischen dem 5% und dem 95% Quantil berechnen, reicht dieser bei positiv wertenden Äußerungen von Männern über ein Intervall von 16,7 Halbtönen und von Frauen über 14,7 Halbtöne. Bei negativ wertenden Aussagen geht der Abstand bei Männern über 11,7 Halbtöne, bei Frauen über 9,1 Halbtöne. Diese Ergebnisse zeigen große Übereinstimmung mit Paeschke (2003), die für neutrale

Äußerungen ein Intervall von 9 Halbtönen errechnet, für die Emotionen Ekel und Langeweile 10 bis 11 Halbtöne und für die Emotion Freude etwa 17 Halbtöne. Enterlein/Bartels/Sendlmeier (2005:25) weisen daher darauf hin, dass der Range der positiv wertenden Äußerungen mit der Emotion Freude vergleichbar ist, die negativen Äußerungen hingegen mit Langeweile, Ekel, Abscheu und Unlust.

In der bereits zitierten Studie zur Sprechweise von Nachrichtensprechern ermitteln Schubert/ Sendlmeier (2005) für den Range sehr niedrige Werte zwischen 3 und 7,1 Halbtönen, was jedoch qualitativ nicht bewertet wird.

Bei den Werbespots wurden mit dem Interquantilsabstand (10-90%) Werte zwischen 8,7 bis 17,7 Halbtönen errechnet. (Siehe Tabelle 10.) Auch unter Berücksichtigung der unterschiedlichen Berechnungsweise für die Spannweite der Grundfrequenz, ist deutlich erkennbar, dass sich nur zwei Spots eindeutig im Bereich des Ausdrucks von positiven Wertungen befinden. Bei den Männerstimmen ist dies Neuroth mit 17,7 und bei den Frauenstimmen Möbelix mit 13,1 Halbtönen. Der Spot Sensodyne liegt mit 8,7 Halbtönen im Bereich neutraler Äußerungen. Die Spots Plantur und Alpecin bewegen sich mit 10,6 und 11,6 Halbtönen hingegen im Bereich negativer Äußerungen. Die drei restlichen Spots, Canesten Bifonazol (13,8 Halbtöne), WC-Ente (14,5 Halbtöne) und Canesten Glutrimazol (12,6 Halbtöne) liegen zwischen den oben genannten Maßzahlen und sind keiner Ausdrucksform zuordenbar.

Die Standardabweichung kann als Maßzahl für die Lebendigkeit oder Monotonie einer Stimme beziehungsweise deren Sprechweise verstanden werden. Ein hoher Wert bedeutet dabei ein hohes Maß an Erregung. Für positiv wertende Äußerungen messen Enterlein/Bartels/Sendlmeier (2005: 23) bei Männer- und Frauenstimmen eine Standardabweichung von 3,9 Halbtönen, für negative Wertungen bei Männern 3,3 und bei Frauen 2,7 Halbtöne.

Bei den Werbespots bewegen sich Plantur mit 4,0 und Sensodyne mit 4,1 Halbtönen im Bereich positiv wertender Äußerungen. Alle anderen Spots liegen deutlich darüber, die Spots Canesten Bifonazol (8,8 Halbtöne), Neuroth (8,9 Halbtöne) und WC-Ente (7,9 Halbtöne) sogar rund doppelt so viel. (Siehe Tabelle 10.) Im Bereich der Standardabweichung kann somit für diese Spots ein hohes bis sehr hohes Maß an Erregung festgestellt werden.

In den einleitenden theoretischen Kapiteln dieser Arbeit wurde bereits auf die hohe Komplexität der Analyse von Stimmen und Sprechweisen hingewiesen. Auch Paeschke (2003:49) betont dazu: „Für das Entstehen einer spezifischen Wirkung ist immer das komplexe Zusammenwirken aller prosodischen Merkmale verantwortlich.“ Die Analyse der einzelnen Parameter kann somit nur einen ersten oberflächlichen Eindruck über die Stimmen und Sprechweisen in den Werbespots vermitteln. Im folgenden fünften Abschnitt werden daher die Spots nun einzeln besprochen und die Berechnungsergebnisse im jeweiligen Zusammenhang sowie hinsichtlich des möglichen Ausdrucks von Kompetenz und der Erregung von Aufmerksamkeit interpretiert.

5. Aufmerksamkeits- und Kompetenzsignale in Fernsehwerbespots

Wie in Kapitel 3.1 ausführlicher beschrieben, verfolgt Werbung das grundlegende Ziel, Aufmerksamkeit zu erregen und damit die Zuwendung des Zielpublikums zu erreichen, sowie auch Kompetenz und dadurch Glaubwürdigkeit zu vermitteln. Im folgenden soll nun untersucht werden, ob diese Intentionen in den Stimmen der untersuchten Werbespots als prosodische Merkmale Ausdruck finden. Als akustische Parameter dieser Merkmale wurden die Sprechgeschwindigkeit sowie die Grundfrequenz mit ihren Eigenschaften Mittelwert, Median, Standardabweichung (Varianz), sowie dem Interquantilsabstand (10-90%) untersucht.

In Kapitel 3.3. wurde ausgeführt, dass das Erregen von Aufmerksamkeit und Aktivierung mit Hilfe starker Reize in der Stimmforschung der Dimension der Erregung entspricht, deren bestimmende prosodische Merkmale eine erhöhte Stimmlage, hohe Varianz, großer Interquantilsabstand sowie eine schnelle Sprechgeschwindigkeit sind. Kompetenz wird hingegen vor allem tiefen Stimmen zugeschrieben, jedoch ebenfalls in Verbindung mit schneller Sprechgeschwindigkeit und einer erhöhten Varianz beziehungsweise Standardabweichung.

Im folgenden werden die Werbespots nun jeweils einzeln auf das Vorhandensein entsprechender Signale für Aufmerksamkeit oder Kompetenz hin untersucht. Die Reihenfolge entspricht dabei der Darstellung in Tabelle 10.

Der Spot **Canesten Bifonazol** wird von einer sehr tiefen Männerstimme gesprochen, die mit einem Mittelwert von 112,6 Hz nahe der Untergrenze von Männerstimmen liegt. Der Median liegt mit 89,8 Hz sogar noch um rund 23 Hz tiefer. Der Interquantilsabstand liegt mit 13,8 Halbtönen im mittleren Bereich, die Standardabweichung ist mit 8,8 Halbtönen hoch. Die globale Sprechgeschwindigkeit ist mit 4,03 Silben pro Sekunde hingegen wiederum langsam. Die sehr tiefe Männerstimme sowie die hohe Standardabweichung vermitteln Signale von Kompetenz. Dies wird jedoch von der eher langsamen Sprechgeschwindigkeit nicht unterstützt.

Der Spot **Neuroth** wird von einer Männerstimme gesprochen, deren Stimmhöhe mit einem Mittelwert von 122,7 Hz im Normalbereich liegt. Der Median liegt mit 112,2 Hz rund 10 Hz tiefer. Der Interquantilsabstand ist mit 17,7 Halbtönen sehr hoch und bewegt sich ziemlich genau im Ausdrucksbereich von Erregung, Freude und positiven Wertungen. Auch die Standardabweichung ist mit 8,9 Halbtönen sehr hoch. Nach diesen zwei Maßzahlen könnten Signale positiver Erregung interpretiert werden. Doch da die Stimmhöhe im Vergleich zu anderen eher tief ist, ist der Ausdruck von Kompetenz anzunehmen. Die sehr langsame Sprechgeschwindigkeit von 3,39 Silben pro Sekunde unterstützt weder den einen noch den anderen Ausdruck.

Die Männerstimme des Spots **Alpecin** kann mit einem Mittelwert von 146,6 Hz und einem Median von 147,1 Hz als hoch bezeichnet werden. Auch die globale Sprechgeschwindigkeit ist mit 5,21 Silben pro Sekunde in Relation zu den anderen Werbespots schnell. Diese zwei Werte können

somit als Signale der Erregung interpretiert werden. Der Interquantilsabstand ist jedoch mit 11,6 Halbtönen eher niedrig und liegt im Bereich des Ausdrucks von negativ wertenden Äußerungen. Die Standardabweichung liegt im Vergleich zu den anderen Spots mit 5,6 Halbtönen im unteren Bereich, ist allerdings noch immer höher als der für positive Äußerungen festgestellte Wert von 3,9 Halbtönen. Auch hier können zwar Signale von Erregung festgestellt werden, jedoch hinsichtlich des Gesamtbildes des stimmlichen Ausdrucks besteht keine Eindeutigkeit.

Bei dem Spot **WC-Ente** liegt der Mittelwert der Grundfrequenz bei für eine Männerstimme hohen 155,1 Hz. Doch der Median ist mit 115,6 Hz um fast 40 Hz tiefer. Betrachtet man die Grafik in Abbildung 10 wird deutlich, dass diese sehr hohe Differenz durch einige, vor allem artikulationsbedingte Spitzenwerte entsteht. Die Männerstimme ist daher als weniger hoch als bei dem Spot Alpecin einzustufen. Doch der Interquantilsabstand (14,5 Halbtöne) und die Standardabweichung (7,9 Halbtöne) liegen in den oberen Bereichen. Die Sprechgeschwindigkeit bewegt sich mit 4,48 Silben pro Sekunde in Relation zu den anderen Spots im mittleren Bereich. Die Parameter der Grundfrequenz können als Signale der Erregung interpretiert werden. Doch aufgrund der mittleren Sprechgeschwindigkeit und des tief liegenden Medians ist auch hier das Gesamtbild nicht ganz eindeutig.

Der Spot **Plantur** wird von einer sehr tiefen Frauenstimme gesprochen, die sich mit einem Mittelwert der Grundfrequenz von 163,3 Hz und einem Median von 152,9 Hz nahe der hohen Männerstimmen von Alpecin und WC-Ente bewegt. Der Interquantilsabstand liegt mit 10,6 Halbtönen im Mittelfeld sowie im Bereich negativer Wertungen, die Standardabweichung ist hingegen mit 4,0 Halbtönen niedrig und im Bereich des Ausdrucks positiver Wertungen. Die globale Sprechgeschwindigkeit bewegt sich mit 4,79 Silben pro Sekunde im mittleren Bereich. Bei Berücksichtigung der vier markanten Pausen ist die Sprechgeschwindigkeit jedoch mit 5,57 Silben pro Sekunde um fast eine Sekunde schneller. Trotz der sehr tiefen Frauenstimme und der mittleren bis eher schnelleren Sprechgeschwindigkeit kann aufgrund des niedrigen Wertes für die Standardabweichung im Gesamtbild kein deutlicher Ausdruck von Kompetenz festgestellt werden.

Der Spot **Canesten Glutrimazol** wird ebenfalls von einer tief liegenden Frauenstimme gesprochen. Auch hier liegt der Mittelwert der Grundfrequenz mit 180,5 Hz unter den durchschnittlichen Werten von 210 bis 240 Hz. Der Median ist mit 165,7 Hz für eine Frauenstimme ebenfalls tief. Der Interquantilsabstand ist mit 12,6 Halbtönen eher hoch, die Standardabweichung ist mit einem Wert von 5,7 Halbtönen in Relation zu den anderen Spots hoch. Die globale Sprechgeschwindigkeit liegt mit 4,78 Silben pro Sekunde im mittleren Bereich. Die tiefe Frauenstimme sowie die hohe Standardabweichung können als Signale der Kompetenz interpretiert werden. Wie beim Spot Neuroth würden auch hier die hohen Werte bei Interquantilsabstand und Standardabweichung genauso gut als Signale der Erregung interpretiert werden können. Doch die tiefe Frauenstimme ist ein deutliches Indiz gegen diesen Ausdruck. Die Sprechgeschwindigkeit liegt im mittleren Bereich und hat daher keine starke Aussagekraft.

Die Stimme im Spot **Sensodyne** liegt mit einem Mittelwert der Grundfrequenz von 221,4 Hz und einem Median von 209,6 Hz im oberen, allerdings normalen Bereich der Frauenstimmen. Der Interquantilsabstand ist mit einem, im Vergleich zu den anderen Spots, niedrigen Wert von 8,7 Halbtönen im Bereich neutraler Äußerungen. Die mit 4,1 Halbtönen ebenfalls nicht sehr hohe Standardabweichung signalisiert hingegen eine positiv wertende Äußerung. Die globale Sprechgeschwindigkeit ist mit 5,34 Silben pro Sekunden eher hoch. Das Gesamtbild dieses Spots kann weder in Richtung des Ausdrucks von Erregung noch in Richtung Kompetenz interpretiert werden.

Der Spot **Möbelix** wird von einer Comic-Stimme gesprochen, die durch tontechnische Manipulation einer Männerstimme[65] erzeugt wurde. Aufgrund des mit 248,1 Hz hohen Mittelwerts der Grundfrequenz sowie des mit 266,5 Hz sehr hohen Medians wurde dieser Spot den Frauenstimmen zugeordnet. Diese Werte liegen jedoch bereits über den für Frauenstimmen üblichen Maßzahlen. Auch der Interquantilsabstand ist mit 13,1 Halbtönen sehr hoch. Die Standardabweichung ist mit 4,9 Halbtönen in Relation zur Höhe der Stimme ebenfalls eher hoch. Die Sprechgeschwindigkeit liegt mit 5,63 Silben pro Sekunde am nächsten zu der als zu schnell bewerteten Grenze von 5,8 Silben pro Sekunde. Bei diesem Spot sind aufgrund der hohen bis sehr hohen Werte bei allen vier Parametern die Signale für Erregung eindeutig gegeben.

In Tabelle 11 werden nun die Ergebnisse der Signalanalyse hinsichtlich des Ausdrucks von Erregung oder Kompetenz zusammengefasst. Die Zahlen werden dabei in eine geordnete qualitative Skalierung umgewandelt und in die Qualitäten „tief bzw. niedrig bzw. langsam“, „mittel“ oder „hoch bzw. schnell“ eingeteilt. Diese Zuordnungen wurden in den vorhergehenden Beschreibungen der Spots besprochen. Entspricht ein Parameter dem möglichen Ausdruck von Kompetenz oder Erregung, wird ihm der Wert Eins zugeteilt und die entsprechende Zelle grau hinterlegt. Eine hellgraue Hinterlegung entspricht einem Signal für Kompetenz, eine dunkelgraue Hinterlegung einem Signal für Erregung. In den letzten zwei Spalten werden die Einsen aggregiert, also zusammengezählt. So kann dargestellt werden, ob und wie weit der Ausdruck von Erregung oder Kompetenz als prosodisches Merkmal einer Stimme deutlich wird.

Für den Ausdruck von Erregung sind, wie weiter oben ausgeführt, folgende vier Parameter von Relevanz: hoher Mittelwert, hohe Standardabweichung, hoher Interquantilsabstand und schnelle Sprechgeschwindigkeit. Für den Ausdruck von Kompetenz sind dies: tiefer Mittelwert, hohe Standardabweichung und schnelle Sprechgeschwindigkeit. Der Interquantilsabstand wird in dem Zusammenhang nicht genannt. Nachdem jedoch ein tiefer oder mittlerer Wert beim Interquantilsabstand in Kombination mit einem tiefen Mittelwert dem Ausdruck von Kompetenz nicht zu widersprechen scheint, wird dieser Parameter in der Tabelle der Kompetenz zugeordnet. So kann eine Gleichgewichtung hinsichtlich der möglichen vier Parameter für den Ausdruck von Erregung

[65] Die hohe Comic-Stimme entsteht vor allem durch schnelleres Aufzeichnen der Originalaufnahme. Durch verlangsamtes Abspielen des Spots im Programm Praat konnte deutlich eine höher gelegene Männerstimme als Ausgangsstimme identifiziert werden.

erreicht werden. Diese Zuordnung erfolgt bei den Spots Canesten Bifonazol und Plantur. Bei den Spots Neuroth und Canesten Glutrimazol ist der Interquantilsabstand hingegen hoch und wird daher als Parameter für den Ausdruck von Erregung gewertet. Nachdem jedoch gleichzeitig auch der Mittelwert der Grundfrequenz tief ist, wird die ebenfalls hohe Standardabweichung als Parameter für Kompetenz und nicht als Parameter für Erregung gewertet.

Werbespot	**Mittelwert**	**Interquantils-abstand**	**Standard-abweichung**	**Sprech-geschwindigkeit**	**Erregung**	**Kompetenz**
Can. Bi.	tief	mittel	hoch	langsam	0	3
Neuroth	tief	hoch	hoch	langsam	1	2
Alpecin	hoch	niedrig	niedrig	schnell	2	0
WC-Ente	hoch	hoch	hoch	mittel	3	0
Plantur	tief	mittel	niedrig	mittel	0	2
Can. Gl.	tief	hoch	hoch	mittel	1	2
Sensodyne	hoch	niedrig	niedrig	schnell	2	0
Möbelix	hoch	hoch	hoch	schnell	4	0

Tabelle 11: Überblick über die Parameter für Erregung (dunkelgrau) und Kompetenz (hellgrau)
1= keine Aussagekraft, 2= wenig eindeutig, 3= eher deutlicher Ausdruck,
4= eindeutiger Ausdruck

Zusammenfasend kann festgestellt werden, dass bei den meisten Spots die insgesamt eher langsamere Sprechgeschwindigkeit weder den Ausdruck von Erregung noch von Kompetenz deutlich unterstützt. Einzig der Spot Möbelix kann mit vier hohen Parametern eindeutig dem Ausdruck von Erregung zugeordnet werden, ebenso, wenn auch nicht ganz so überzeugend, der Spot WC-Ente aufgrund der hohen Werte bei den Parametern der Grundfrequenz. Der Spot Alpecin tendiert in Richtung Erregung, ist jedoch im Gesamtbild wenig eindeutig. Der Spot Sensodyne hat wie der Spot Alpecin zwei Parameter (Mittelwert und Sprechgeschwindigkeit) für den Ausdruck von Erregung, doch ist der Mittelwert der Grundfrequenz in Relation zu den anderen Stimmen der Spots zwar als hoch einzustufen, jedoch nicht als erhöht im Vergleich zu üblichen Frauenstimmen. Das Gesamtbild des Spots zeigt, vor allem auch aufgrund der niedrigen Werte bei Interquantilsabstand und Standardabweichung, keine klare Tendenz.

Der Spot Canesten Bifonazol hat deutliche Signale der Kompetenz, die allerdings von einer niedrigen Sprechgeschwindigkeit nicht unterstützt werden. Der Spot Plantur kann aufgrund der sehr tiefen Frauenstimme, dem mittleren Interquantilsabstand und der mittleren Sprech-

geschwindigkeit eher dem Ausdruck von Kompetenz zugeordnet werden, der jedoch von einem niedrigen Wert bei der Standardabweichung nicht unterstützt wird. Die Spots Canesten Glutrimazol und Neuroth wiederum zeigen ein ähnliches und eher widersprüchliches Bild. Beide haben einen tiefen Mittelwert der Grundfrequenz, eine hohe Standardabweichung und einen hohen Interquantilsabstand. Es könnten daher auch zwei Parameter für den Ausdruck von Erregung gezählt werden. Ein tiefer Mittelwert der Grundfrequenz spricht jedoch eindeutig gegen den Ausdruck von Erregung und daher wird die Standardabweichung dieser zwei Spots als Parameter der Kompetenz gewertet. Ob ein hoher Interquantilsabstand den Ausdruck von Kompetenz unterstützt oder ihm entgegenarbeitet, kann aus den vorliegenden Forschungsergebnissen nicht abgeleitet werden. In jedem Fall ist auch aufgrund der langsamen bis mittleren Sprechgeschwindigkeit der Ausdruck von Kompetenz nicht sehr überzeugend.

6. Schlussbemerkungen und weiterführende Forschungsfragen

Die forschungsleitenden Fragen der vorliegenden Arbeit waren:

- Können Stimmen in Werbespots identifiziert werden, die anhand bestimmter messbarer physikalischer Parameter und prosodischer Merkmale dem Kommunikationsziel der Erregung von Aufmerksamkeit und Aktivierung entsprechen?
- Können Stimmen in Werbespots identifiziert werden, die messbare Signale der Kompetenz aussenden und die dadurch dem Kommunikationsziel Vertrauen zu erwecken zuordenbar sind?

Die Bearbeitung dieser Fragestellung war erkenntnisoffen, da eine Hypothesenbildung auf Grund fehlender relevanter Forschungsergebnisse in diesem Arbeitsbereich nicht möglich war.

Insgesamt wurden die akustischen Parameter der Stimmen von acht Fernsehwerbespots detailliert analysiert. Dabei konnten insgesamt drei Spots identifiziert werden, die eindeutige Signale der Erregung (Möbelix, WC-Ente) oder Kompetenz (Canesten Bifonazol) aussenden. Die zwei forschungsleitenden Fragen können somit mit einem Ja beantwortet werden. Die anderen fünf Spots waren weniger eindeutig und zum Teil widersprüchlich.

Dazu soll an dieser Stelle noch einmal darauf hingewiesen werden, dass insgesamt vier prosodische Merkmale (Mittelwert, Interquantilsabstand und Standardabweichung der Grundfrequenz sowie Sprechgeschwindigkeit) der Stimmen analysiert wurden, deren Auswahl und Bewertung auf der Basis vorhandener Forschungsliteratur erfolgte. In Relation zur Komplexität der Signale von Stimmen und der Vielzahl an Ausdrucksmöglichkeiten erscheinen diese vier Parameter als wenig. Jedoch konnten sie in der bisherigen Forschung als für den vorliegenden Kontext relevant identifiziert werden. Andere Parameter - in dem Zusammenhang seien vor allem die Rauigkeit einer Stimme wie auch die Anzahl, Stärke und Charakteristik der Formanten erwähnt – müssen unbeachtet bleiben, da ihr Einfluss auf spezifische Wirkungen von Stimmen und Sprechweisen bislang zu wenig erforscht werden konnte und wurde.

Um die Ergebnisse der Signalanalyse zu ergänzen und weitergehende Aussagen zur Wirkung der untersuchten Werbespots hinsichtlich des Ausdrucks von Kompetenz oder Erregung machen zu können, ist die außergewöhnliche Wahrnehmungsfähigkeit des menschlichen Hörorgans unerlässlich. Denn nur mit Hilfe der auditiven Perzeption kann die Charakteristik von Stimmen und Sprechweisen umfassend interpretiert werden. Der nächste Arbeitsschritt wäre daher, diese Spots der auditiven Rezeption und Bewertung zu unterziehen und folgende weiterführende Forschungsfrage zu formulieren:

- Entspricht das objektiv Gemessene, also die akustischen Signale der Stimmen, repräsentiert durch die entsprechenden Signalparameter für einerseits Erregung sowie andererseits Kompetenz, auch der kognitiv-emotionalen Wahrnehmung durch die Rezipienten?

Insbesondere bei den Spots Canesten Glutrimazol und Neuroth, bei denen widersprüchliche Merkmale aus den Signalen abgeleitet wurden, könnte die auditive Perzeption und Bewertung durch eine statistisch ausreichend große Probandengruppe weiteren Erkenntnisgewinn hinsichtlich der Rolle der prosodischen Merkmale und deren Wirkung bringen.

Für die Größe und Zusammensetzung einer Probandengruppe wären mehrere Aspekte von Relevanz: In Kapitel 2.2. wurde darauf hingewiesen, dass der Prozess des Hörens subjektiv und von physiologischen Voraussetzungen abhängig ist. Im Rahmen einer Befragung zur auditiven Perzeption der Werbespots müsste daher auch die Toleranz gegenüber akustischen Erregungsreizen überprüft werden. Dies könnte mit Hilfe der Bewertung von nonverbalen Hörbeispielen, die entsprechende Signalparameter aufweisen, durchgeführt werden. Ein wesentlicher Faktor wäre in diesem Zusammenhang auch die Altersstruktur der Probandengruppe. Denn bei den meisten Menschen verändern sich mit fortschreitendem Alter die auditiven Fähigkeiten. Es entwickelt sich häufig Schwerhörigkeit. Unter anderem nimmt auch das Hörvermögen hoher Frequenzen messbar ab.[66] Diesbezüglich ist zu fragen, ob diese physiologischen Veränderungen auch die Wahrnehmung und Beurteilung von Stimmen beeinflussen, ob zum Beispiel hohe Stimmen als weniger hoch und daher weniger Erregung auslösend wahrgenommen werden oder ob andererseits tiefe Stimmen schlechter verstanden und daher entgegen der Intention Kompetenz und Vertrauen zu vermitteln eher Verunsicherung oder Ablehnung auslösen. Es kann dazu somit eine zweite weiterführende Forschungsfrage formuliert werden:

- Gibt es bei der auditiven Perzeption der ausgewählten Werbespots altersspezifische Unterschiede bei der Wahrnehmung und Beurteilung der gehörten Stimmen?

Doch nicht nur die auditiven Fähigkeiten, auch andere Hörgewohnheiten und Erwartungshaltungen könnten ältere Menschen von jüngeren unterscheiden. Ein diesbezüglicher Erkenntnisgewinn würde wiederum von Interesse für zielgruppengerechte Werbearbeit sein.

In diesem Zusammenhang wäre insbesondere die auditive Bewertung des Spots Neuroth interessant, der durch eine sehr langsame Sprechgeschwindigkeit auffällt und dessen Signalparameter in Summe weder dem Ausdruck von Erregung noch von Kompetenz zugeordnet werden können. Dieser Spot wirbt für Hörgeräte und wendet sich offensichtlich an ältere beziehungsweise

---

[66] Hörstörungen können angeboren sein oder auch durch Entzündungen, Infektionen, Medikamente, Lärmbelastungen oder akute Schalltraumata (Knallen) erworben werden. (Dazu findet sich eine gute Übersicht bei Schmidt 2011:5f.)
Die Schwerhörigkeit im Alter (Presbyakusis) hat jedoch bislang keine eindeutig bekannte Ursache: „Schwerhörigkeit tritt im Alter statistisch gehäuft auf, aber es gibt keine ‚natürliche' Altersschwerhörigkeit. Dabei scheinen medizinische Risikofaktoren für das Entstehen einer Schwerhörigkeit weniger wichtig zu sein als genetische Faktoren und vor allem umweltbedingte Einflüsse, beispielsweise Lärm. In den letzten Jahrzehnten hat sich das Hörvermögen alter Menschen in den tiefen und mittleren Frequenzen im Durchschnitt verschlechtert, ist aber in den hohen Frequenzen weitgehend gleich geblieben. Die meisten Schwerhörigkeiten im Alter entstehen sowohl durch Veränderungen der Haarzellen des Innenohres als auch durch degenerative Abbauprozesse der zentralen Hörbahn." (Hesse/Laubert 2005:A2864)

hörgeschädigte Menschen. Die Ergebnisse einer auditiven Perzeption könnten gerade hier sehr aufschlussreich sein.

In der für den vorliegenden Kontext relevanten Forschungsliteratur werden geschlechtsspezifische Aspekte im Zusammenhang mit der auditiven Rezeption von Stimmen nicht thematisiert. Bei einer ausreichend großen Gruppe von Versuchspersonen könnte daher auch untersucht werden, ob es diesbezüglich Unterschiede in der Bewertung der Werbespots gibt. Dazu kann folgende weiterführende Forschungsfrage formuliert werden:

- Gibt es bei der auditiven Perzeption der ausgewählten Werbespots geschlechtsspezifische Unterschiede bei der Wahrnehmung und Beurteilung der gehörten Stimmen?

Die drei weiterführenden Forschungsfragen betreffen Aspekte der auditiven Perzeption. Diesbezüglich soll abschließend noch einmal betont werden, dass gerade die Kombination der zwei Forschungsmethoden, also der Signalanalyse einerseits und der auditiven Perzeption andererseits, weiteren und tiefer gehenden Erkenntnisgewinn hinsichtlich des stimmlichen Ausdrucks und der Wirkung von gesprochenen Botschaften bringen würde. Erst durch die Verbindung der Messung des Signals mit der auditiven Wahrnehmung und Bewertung desselben können auf objektiven Daten basierende Interpretationen der Bedeutung und Wirkung von prosodischen Merkmalen gesprochener Botschaften erfolgen. Ergebnisse dieser Forschung wären im Zusammenhang mit rezipientenorientierter vokaler Kommunikation, wie in den audiovisuellen Medien oder der Werbung in denselben, von Interesse. Durch die technologischen Entwicklungen in den letzten Jahren und den damit verbundenen Fortschritten bei der stimmenspezifischen signalanalytischen Software unter Einsatz neuer Algorithmen wurden auch im Bereich der Stimm- und Kommunikationsforschung neue Arbeitsbereiche eröffnet, die es in Zukunft zu nutzen gilt.

7. ZUSAMMENFASSUNG

Forschungsgegenstand dieser Arbeit ist die menschliche sprachliche Kommunikation in audio-visuellen Medien. Konkret geht es um die Frage, ob kommunikative Ziele der Werbung als prosodische Merkmale von Sprechstimmen identifiziert werden können.

Die forschungsleitenden Fragen der vorliegenden Arbeit sind:

- Können Stimmen in Werbespots identifiziert werden, die anhand bestimmter messbarer physikalischer Parameter und prosodischer Merkmale dem Kommunikationsziel der Erregung von Aufmerksamkeit und Aktivierung entsprechen?
- Können Stimmen in Werbespots identifiziert werden, die messbare Signale der Kompetenz aussenden und die dadurch dem Kommunikationsziel Vertrauen zu erwecken zuordenbar sind

Die Bearbeitung dieser Fragestellungen war erkenntnisoffen, da eine Hypothesenbildung auf Grund fehlender relevanter Forschungsergebnisse in diesem Arbeitsbereich nicht möglich war.

Zu Beginn der Arbeit werden die wissenschaftlichen und theoretischen Grundlagen besprochen. Mit dem Ziel, den nicht verhinderbaren Einfluss des Körpers sowie emotionaler Zustände auf die Stimme zu verdeutlichen, werden die physiologischen Grundlagen des Stimmapparates erklärt. Denn in Verbindung mit der Sprechweise gibt der unverwechselbare und einzigartige Stimmcharakter jedes Menschen der gesprochenen Sprache eine Vielfalt an Ausdrucksmöglichkeiten, die weit über diejenigen des geschriebenen Wortes hinausgehen. Bei der Betrachtung der Physiologie des Hörapparates wird dargelegt, dass die menschliche auditive Wahrnehmungs- und Interpretationsfähigkeit dieser Vielfalt in jedem Fall gerecht zu werden vermag. Im Anschluss daran werden die Zusammenhänge und Wechselwirkungen von Stimme und Sprache, von Stimmeigenschaften und Sprechweise, der Prosodie, aufgezeigt.

Die mathematische Theorie der Kommunikation nach Shannon (1949) wird als Grundlage gewählt um die gestellten Forschungsfragen zu bearbeiten. Denn basierend auf dem Prozess der Signalübertragung werden darin auch die Eigenschaften von Kommunikator (information source), Rezipient (receiver), Medium (transmitter) und Botschaft (signal) behandelt sowie Kodierung und Dekodierung der Signale und mögliche Störungen dieses Prozesses (noise source) miteinbezogen. Diese detaillierte Betrachtungsweise des kommunikativen Prozesses wird der Komplexität der menschlichen sprachlichen Kommunikation am ehesten gerecht.

Wesentliche strategische kommunikative Ziele der Werbung sind die Erregung von Aufmerksamkeit und die Vermittlung von Kompetenz. Basierend auf bisherigen Ergebnissen der Stimmforschung können folgende, in dem Zusammenhang relevante, akustische Parameter identifiziert werden: die Sprechgeschwindigkeit sowie die Grundfrequenz der Stimme mit ihren Eigenschaften Mittelwert (Stimmhöhe), Range (Spannweite) und Varianz (Lebendigkeit). Aufmerksamkeit und Aktivierung durch starke Reize zu erzeugen entspricht hier der Dimension der Erregung. Deren

prosodische Merkmale sind erhöhte Stimmlage, hohe Varianz, großer Range und eine schnelle Sprechgeschwindigkeit. Erhöhte Varianz und schnelle Sprechgeschwindigkeit sind auch Merkmale, die in der auditiven Perzeption mit Kompetenz in Verbindung gebracht werden. Der bestimmende Faktor und wesentliche Unterschied zum Ausdruck von Erregung ist jedoch die Höhe der Grundfrequenz. Denn Kompetenz wird in unserer Sprachkultur vor allem tiefen Männerstimmen (geringer Mittelwert der Grundfrequenz) zugeordnet, die auch mit Vertrauenswürdigkeit und Souveränität in Verbindung gebracht werden. Somit können Stimmen nur eines der beiden strategischen kommunikativen Ziele der Werbung ausdrücken: entweder Aufmerksamkeit und Aktivierung oder Kompetenz.

In der Signalanalyse von Fernsehwerbespots wird untersucht, ob prosodische Merkmale von Stimmen und Sprechweisen identifiziert werden können, die einem der genannten Ziele zuordenbar sind. Nach der Beschreibung des Auswahlprozesses von zur Stimmanalyse geeigneten Fernsehwerbespots werden die Berechnungen der akustischen Parameter Sprechgeschwindigkeit sowie Grundfrequenz mit ihren Eigenschaften Mittelwert und Median, Standardabweichung (Varianz) sowie dem Range beziehungsweise Interquantilsabstand (10-90%) der verwendeten Stimmen dargestellt und erläutert.

Insgesamt werden die akustischen Parameter der Stimmen von acht Fernsehwerbespots analysiert. Dabei können drei Spots identifiziert werden, die eindeutige Signale der Erregung (Möbelix, WC-Ente) oder der Kompetenz (Canesten Bifonazol) aussenden. Die zwei forschungsleitenden Fragen können somit mit einem Ja beantwortet werden. Die anderen fünf Spots lassen sich aufgrund ihrer wenig eindeutigen oder auch widersprüchlichen prosodischen Merkmale keinem der beiden strategischen kommunikativen Ziele der Werbung zuordnen.

Um über diesen Befund der Signalanalyse hinausgehende Aussagen machen zu können, bedarf es der außergewöhnlichen Wahrnehmungsfähigkeit des menschlichen Hörorgans. Ein nächster Arbeitsschritt wäre daher, die signalanalytisch untersuchten Werbespots der auditiven Rezeption und Bewertung zu unterziehen. Weiterführende Forschungsfragen könnten wie folgt formuliert werden:

- Entspricht das objektiv Gemessene, also die akustischen Signale der Stimmen, repräsentiert durch die entsprechenden Signalparameter für einerseits Erregung sowie andererseits Kompetenz, auch der kognitiv-emotionalen Wahrnehmung durch die Rezipienten?
- Gibt es bei der auditiven Perzeption der ausgewählten Werbespots altersspezifische Unterschiede bei der Wahrnehmung und Beurteilung der gehörten Stimmen?
- Gibt es bei der auditiven Perzeption der ausgewählten Werbespots geschlechtsspezifische Unterschiede bei der Wahrnehmung und Beurteilung der gehörten Stimmen?

Ergebnisse dieser Forschung könnten im Zusammenhang mit rezipientenorientierter vokaler Kommunikation, wie in den audiovisuellen Medien oder der Werbung in denselben, von Interesse sein.

Literaturverzeichnis

ALLHOF, Dieter-W. (Hrsg., 1983): Sprechpädagogik - Sprechtherapie. (Beiträge zur Sprechwissenschaft und Sprecherziehung, Sprache und Sprechen, Bd. 2) Frankfurt am Main.

BADURA, Bernhard (2004): Mathematische und soziologische Theorie der Kommunikation. In: Burkart, Roland/Hömberg, Walter (Hrsg.): Kommunikationstheorien. Wien. S. 16-23.

BERENDT, Joachim-Ernst (1998): Ich höre, also bin ich. In: Vogel, Thomas (Hrsg.): Über das Hören: einem Phänomen auf der Spur. 2. bearb. Auflage, Tübingen. S. 69-90.

BURKART, Roland (2002): Kommunikationswissenschaft. 4. Aufl. Wien/Köln/Weimar.

CHAIKA, Elaine (1989): Language. The Social Mirror. Cambridge, NY.

DITTMANN, Roland (1994): Entwicklung eines Expertensystems zur Beurteilung von Radiowerbung. Dissertation an der Universität des Saarlandes, Rechts- und Wirtschaftswissenschaftliche Fakultät. Saarbrücken.

ECKERT, Hartwig/LAVER, John (1994): Menschen und ihre Stimmen. Aspekte der vokalen Kommunikation. Weinheim.

EISINGER, Günther (2002): Stimmfrequenzmessung unter physischer und psychischer Belastung zur Beurteilung emotionalen Stresses. Diplomarbeit zur Erlangung des Magistergrades der Naturwissenschaften an der Fakultät für Human- und Sozialwissenschaften der Universität Wien. Wien.

ENTERLEIN, Ines/BARTELS, Astrid/SENDLMEIER, Walter (2005): Prosodische Indikatoren der Sprechereinstellung. In: Sendlmeier, Walter/Bartels, Astrid (Hrsg.): Stimmlicher Ausdruck in der Alltagskommunikation. Reihe Mündliche Kommunikation, Bd. 4. Berlin. S. 9-38.

FAHRMEIR, Ludwig / KÜNSTLER, Rita / PIGEOT, Iris / TUTZ, Gerhard (2011): Statistik. Der Weg zur Datenanalyse. 7. Auflage, korrigierter Nachdruck. Heidelberg/Dordrecht/London/New York.

FELSER, Georg (2007): Werbe- und Konsumentenpsychologie. Nachdruck 2011 der 3. Aufl. 2007. Berlin/Heidelberg.

FIUKOWSKI, Heinz (1999): Zur Präsentation von Nachrichten im Hörfunk. Ein Arbeits- und Erfahrungsbericht. In: Krech, Eva-Maria/Stock, Eberhard (Hrsg.): Sprechwissenschaft – Zu Geschichte und Gegenwart. Hallesche Schriften zur Sprechwissenschaft und Phonetik, Bd. 3, S. 145-156. Frankfurt am Main.

GEISSNER, Hellmut K. (1981): Sprechwissenschaft. Theorie der mündlichen Kommunikation. (Monographien Literatur+Sprache+Didaktik, Bd. 26) Königstein/Ts.

GEISSNER, Hellmut K. (Hrsg.) (2004): Das Phänomen Stimme in Kunst, Wissenschaft, Wirtschaft. 4. Stuttgarter Stimmtage 2002. St. Ingbert.

GOLDSTEIN, E. BRUCE (2002): Wahrnehmungspsychologie. (2. dt. Aufl.; Hrsg: Ritter, Manfred) Heidelberg/Berlin.

GRAUMANN, Carl Friedrich (1972): Interaktion und Kommunikation. In: ders. (Hrsg.): Handbuch der Psychologie. Bd. 7: Sozialpsychologie. Göttingen.

HAUSER, Marc D. / CHOMSKY, Noam / FITCH, Tecumseh W. (2002): The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? In: SCIENCE, Vol. 298, S.1569-1579.

HASELOW, Alexander (2015): Final particles in spoken German. In: Hancil, Sylvie / Haselow, Alexander / Post, Margje (Hg.): Final Particles. Berlin/Boston.

HESSE, Gerhard / LAUBERT, Armin (2005): Hörminderung im Alter - Ausprägung und Lokalisation. In: Deutsches Ärzteblatt, Jg. 102, Heft 42 / 21. Oktober 2005, S. A 2864-A 2868.

HILTENSPERGER, Christina (2004): Untersuchung der Lautstärkeempfindung von Schwerhörigen mit der „Methode der Linienlänge“. Dissertation, Medizinische Fakultät der LMU, München.

KLINGEBIEL, Randolf (2002): Evaluation neuer radiologischer Bildgebungstechniken in der otologischen Diagnostik. Habilitationsschrift, Berlin.

KROEBER-RIEL, Werner / ESCH, Franz-Rudolf (2011): Strategie und Technik der Werbung. Verhaltens- und neurowissenschaftliche Erkenntnisse. 7. aktual. und überarb. Auflage. Stuttgart.

LENKE, Nils / LUTZ, Hans-Dieter / SPRENGER, Michael (1995): Grundlagen sprachlicher Kommunikation. München.

LEONHARDT, Helmut (1973): Innere Organe. dtv-Atlas der Anatomie Bd. 2. Stuttgart/München.

MANASSI, Sabina (2003): Pädagogik des Horchens. Eine Einführung. In: Tomatis, Alfred A.: Der Klang des Lebens. Vorgeburtliche Kommunikation - die Anfänge der seelischen Entwicklung. 12. Auflage, Hamburg. S. 9-34.

MAYER, Jörg (2010): Linguistische Diagnostik. Apparative phonetische Methoden: Elektrolaryngographie (ELG) bzw. Elektroglottographie (EGG), in: Sprache und Gehirn. Ein neurolinguistisches Tutorial. http://www2.ims.uni-stuttgart.de/sgtutorial/elg.html (22.9.2014)

MAYR, Nora (2006): Stimmen in der Radiowerbung. Dipl. Arbeit. Wien.

MÜLLER, Gerhard (2009): Werbung und Vertrauen - Widerspruch oder Notwendigkeit? Diplomarbeit am Institut für Publizistik- und Kommunikationswissenschaften der Universität Wien.

NAWRATIL, Ute (2006): Glaubwürdigkeit in der sozialen Kommunikation. 2. Auflage, München. Digitale Ausgabe: http://epub.ub.uni-muenchen.de/archive/00000941/ (18.7.2014).

NÖTH, Winfried (2000): Handbuch der Semiotik. 2. Vollständig neu bearb. und erw. Auflage. Stuttgart/Weimar.

PAESCHKE, Astrid (2003): Prosodische Analyse emotionaler Sprechweise. Reihe Mündliche Kommunikation, Bd. 1. Berlin.

PFITZINGER, Hartmut R. (2001): Phonetische Analyse der Sprechgeschwindigkeit. Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM) Nr. 38, S.117-264.

PROSS, Harry (1972): Medienforschung. Darmstadt.

SAUSSURE, Henry de (1931/1967): Grundfragen der allgemeinen Sprachwissenschaft. Hrsg. v. Bally, Charles / Sechehaye, Albert. Berlin.

SCHERER, Klaus R. (Hrsg., 1982): Vokale Kommunikation. Nonverbale Aspekte des Sprachverhaltens. Weinheim, Basel.

SCHMIDT, Claus-Michael (2011): Funktionelle Anatomie des Gehörs. Skriptum, Universität Münster.
http://www.klinikum.uni-muenster.de/fileadmin/ukminternet/daten/zentralauftritt/forschung-lehre/schulen/logopaedie/Script_Paedaudio_2011.pdf (31.8.2013).

SCHUBERT, Antje / SENDLMEIER, Walter (2005): Was kennzeichnet gute Nachrichtensprecher im Hörfunk ? Eine perzeptive und akustische Analyse von Stimme und Sprechweise. In: SENDLMEIER, Walter (Hrsg.): Sprechwirkung - Sprechstile in Funk und Fernsehen. Mündliche Kommunikation Bd. 3. Berlin.

SCHWEIGER, Günter / Schrattenecker, Gertraud (2001): Werbung. 5. neu bearb. Auflage. Stuttgart.

SEIMER, Andreas (2006): Aspekte der Hör- und Sprachentwicklung. Grundlagen geglückter Kommunikation, in: Universitas, 61. Jg. Nr. 3 /2006, S. 267-286, Heidelberg.

SENDLMEIER, Walter (2005): Mündlichkeit - Sprechstile in den Medien. In: SENDLMEIER, Walter (Hrsg.): Sprechwirkung - Sprechstile in Funk und Fernsehen. Mündliche Kommunikation Bd. 3. Berlin.

SENDLMEIER, Walter/BARTELS, Astrid (Hrsg., 2005): Stimmlicher Ausdruck in der Alltagskommunikation. Reihe Mündliche Kommunikation, Bd. 4. Berlin.

SHANNON, Claude E. / Weaver, Warren (1949, 1998): The Mathematical Theory of Communication. Urbana and Chicago.

STANDKE, Reiner (1993): Methoden der digitalen Sprachverarbeitung in der vokalen Kommunikationsforschung. (Europäische Hochschulschriften, Reihe VI Psychologie, Bd. 402) Frankfurt am Main et al.

STIER, Winfried (1999): Empirische Forschungsmethoden. 2. verb. Auflage. Berlin/Heidelberg.

STRASSNER, Erich (1982): Fernsehnachrichten: eine Produktions-, Produkt- und Rezeptionsanalyse. Tübingen

TERHARDT, Ernst (1998): Akustische Kommunikation. Grundlagen mit Hörbeispielen. Berlin/Heidelberg.

TOMATIS, Alfred A. (1981, 1987, 2003): Der Klang des Lebens. Vorgeburtliche Kommunikation - die Anfänge der seelischen Entwicklung. Franz. Originalausgabe (1981), Paris, deutsch (1987) und 12. Auflage (2003), Hamburg.

TROJAN, Felix (1975): Biophonetik. (Hrsg.: Schendl, Herbert). Mannheim/Wien/Zürich.

WERLEN, Iwar (1984): Ritual und Sprache. Zum Verhältnis von Sprechen und Handeln in Ritualen. Tübingen.

WESTPHAL, Kristin (2002): Wirklichkeiten von Stimmen. Grundlegung der Theorie der medialen Erfahrung. Habilitationsschrift. Gießen.

WILLIAMS, Carl E./STEVENS, Kenneth N. (1981): Vocal correlates of emtotional states. In: Darby, J. (Ed.) (1981): Speech Evaluation in Psychiatry. New York, S. 221-240. Dt. Übersetzung in: SCHERER (1982), S. 307-325.

WINKLER, Ralf (2008): Merkmale junger und alter Stimmen: Analyse ausgewählter Parameter im Kontext von Wahrnehmung und Klassifikation. Mündliche Kommunikation Bd.6. Berlin.

ZWICKER, Eberhard / FASTL, Hugo (1999): Psychoacoustics. Facts and Models. Zweite überarbeitete Auflage, Heidelberg/New York.

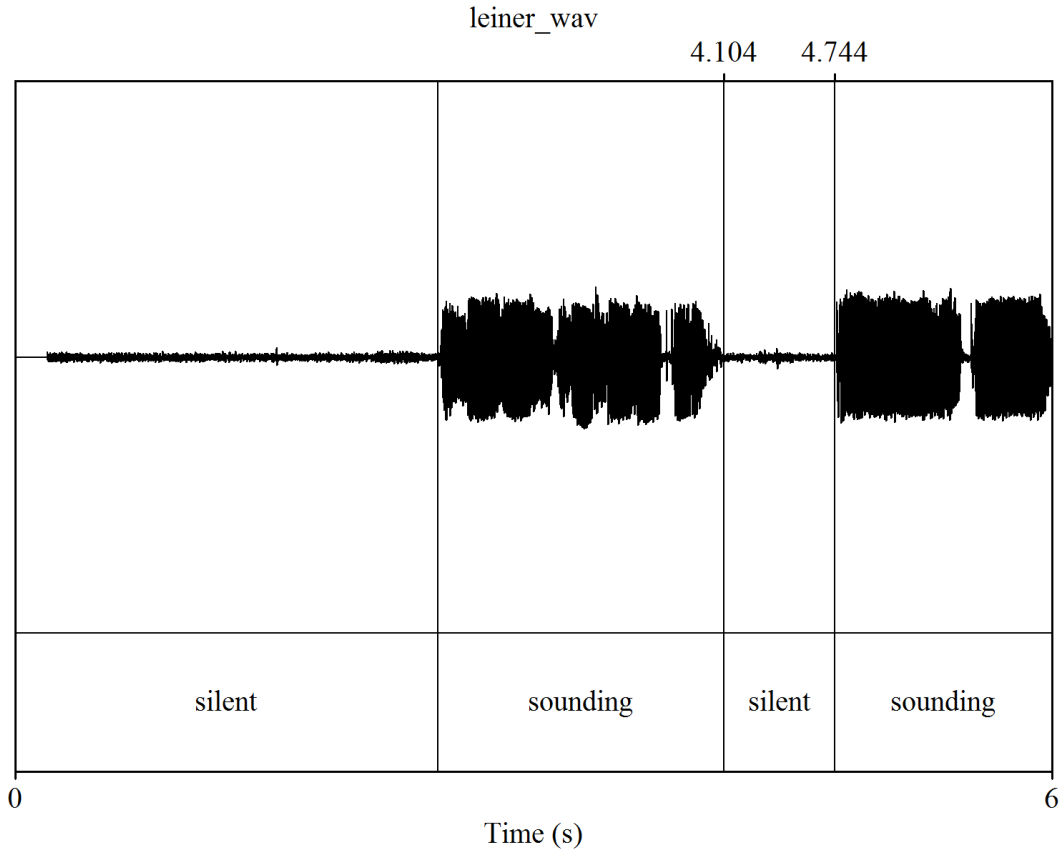
Praat: Das Sprachanalyseprogramm Praat wurde entwickelt von Boersma, Paul und Weenink, David, Institute of Phonetics Sciences, University of Amsterdam. http://www.praat.org oder http://www.fon.hum.uva.nl/praat/ (16.8.2015)

ANHANG 1: AMPLITUDENVERLÄUFE, SPRECHGESCHWINDIGKEIT UND DOKUMENTATION DER WERBESPOTS

Die Werbespots sind im folgenden nach der Länge der Sprechdauer aufsteigend gereiht. Die Grafiken der Amplitudenverläufe sind mit dem Programm Praat erzeugt. Die Schwellenwerte zur Lautheit (Schalldruckpegel) wurden bei jedem Spot händisch und mit auditiver Kontrolle angepasst, um die Stimmen von etwaigen Hintergrundgeräuschen trennen zu können. Sprechpausen und Sprechdauer wurden ebenfalls bei Bedarf modifiziert. Diese Arbeitsschritte werden bei den jeweiligen Spots im Detail erklärt. Pausen werden ab einer Länge von 0,25 Sekunden ausgewiesen und berechnet. Die Dokumentation der gesprochenen Texte und der Sprechpausen sowie die Berechnung der Sprechgeschwindigkeit erfolgten händisch.

Dokumentiert werden: die Gesamtdauer der Aufnahme, die Sprechdauer in Sekunden mit zwei Nachkommastellen (in Klammer gesetzt ist die Anfangs- und Endzeit der Sprechdauer), die Differenz der Sprechdauer in Prozent und das jeweilige Ergebnis für die Sprechgeschwindigkeit, gerundet auf zwei Nachkommastellen. Pausen werden mit ihrer Länge in Sekunden (sec) auf zwei Nachkommastellen gerundet und an den entsprechenden Stellen der Texte in eckige Klammern gesetzt. Um einen möglichen Einfluss der Pausen auf die Sprechgeschwindigkeit deutlich machen zu können, wird diese zweimal berechnet: einmal mit der Gesamtdauer (global), einmal unter Abzug der Pausenzeiten. Angeführt ist ebenfalls die Differenz dieser zwei Ergebnisse.

SPOT: LEINER



Text: „Service zum Verlieben. [4,10-4,74] Der Leiner ist meiner."

Dauer der Aufnahme: 6 sec, Sprechdauer gesamt: 3,55 sec (2,45 – 6,00)

Pausen: 1 Pause mit einer Gesamtlänge von 0,64 sec

Sprechdauer minus Pausen: 2,91 sec

Differenz Sprechdauer: 18,03%

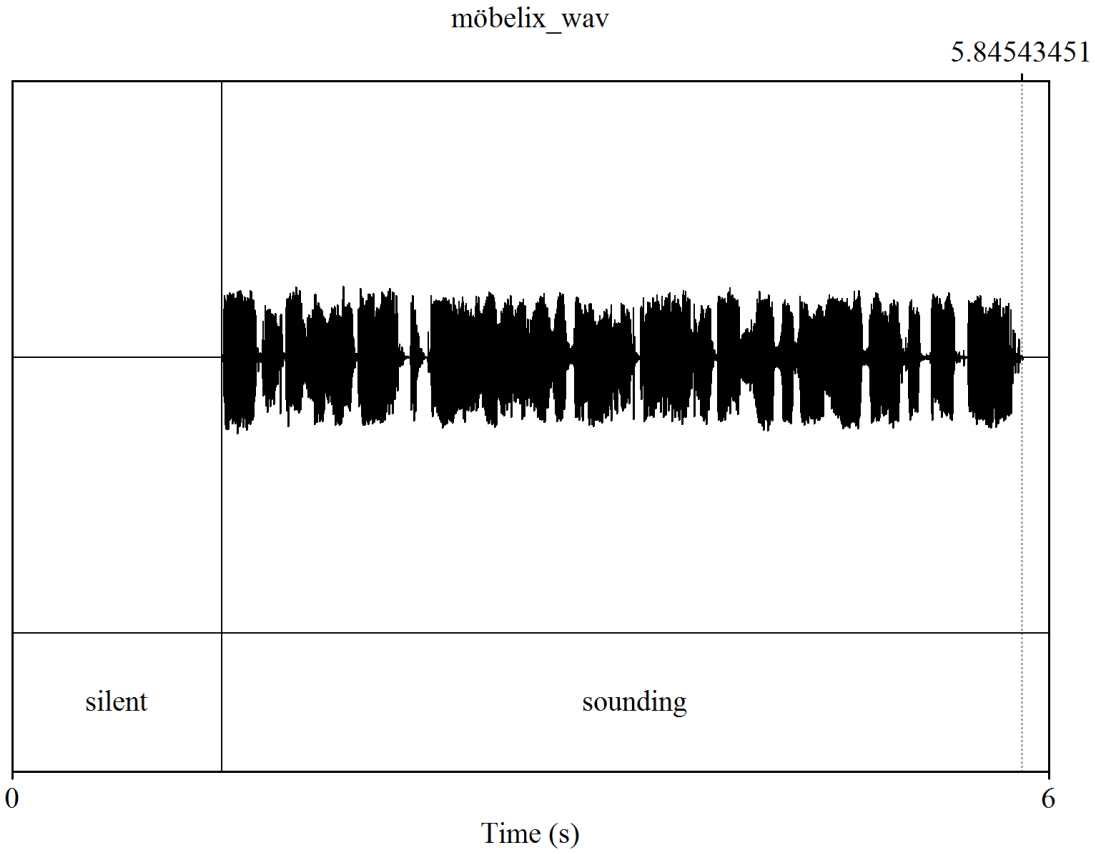
Sprechgeschwindigkeit global: 12 Silben in 3,55 sec = 3,38 Silben /sec

Sprechgeschwindigkeit minus Pausen: 12 Silben in 2,91 sec = 4,12 Silben /sec

Differenz Silben: 0,74 Silben/sec

Die leichten Hintergrundgeräusche sind im Amplitudenverlauf sichtbar. Sie werden aber als silent, also als Pause berechnet, da sie unter dem Schwellenwert von -25 dB Differenz zur maximalen Lautheit liegen. Durch die Vorgabe dieses Differenzwertes kann die Stimme berechnet werden, ohne dass durch die Hintergrundgeräusche die Ergebnisse beeinflusst werden.

SPOT: MÖBELIX



Text: „Jetzt auch im Internet. Massenweise Jubiläumspreise unter www möbelix at.“

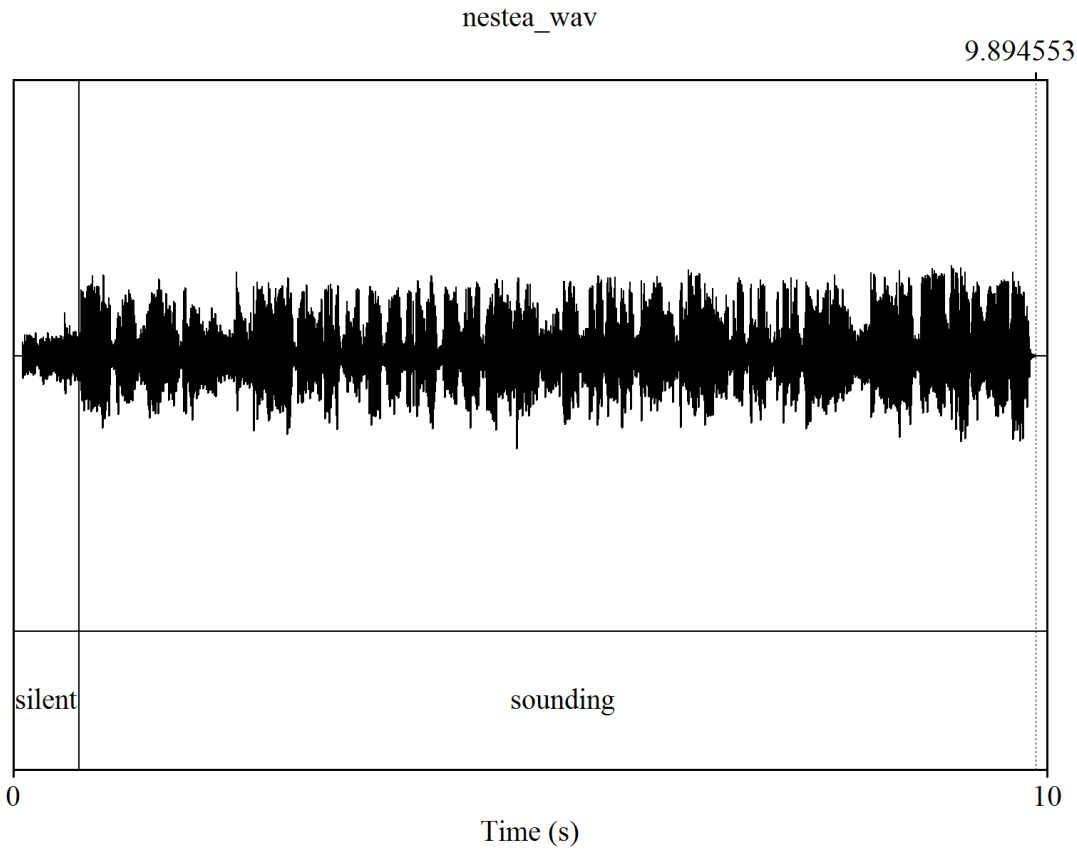
Dauer der Aufnahme: 6 sec, Sprechdauer gesamt: 4,62 sec (1,22 – 5,84)

Pausen: keine

Sprechgeschwindigkeit global: 26 Silben in 4,62 sec = 5,63 Silben /sec.

Bei der Grafik ist eine Diskrepanz in der Länge des gesprochenen Textes. Da die minimale Pausenlänge mit 0,25 Sekunden definiert ist, wird der Schluss des Spots als Sound angezeigt. Der Cursor zeigt diese Diskrepanz des Messens bei 5,84 Sekunden an.

SPOT: NESTEA



Text: „Nestea Greentea. Dank der Süßkraft aus dem Extrakt der Stevia-Pflanze konnten wir den Zucker um dreißig Prozent reduzieren. Erleb mal was Neues.“

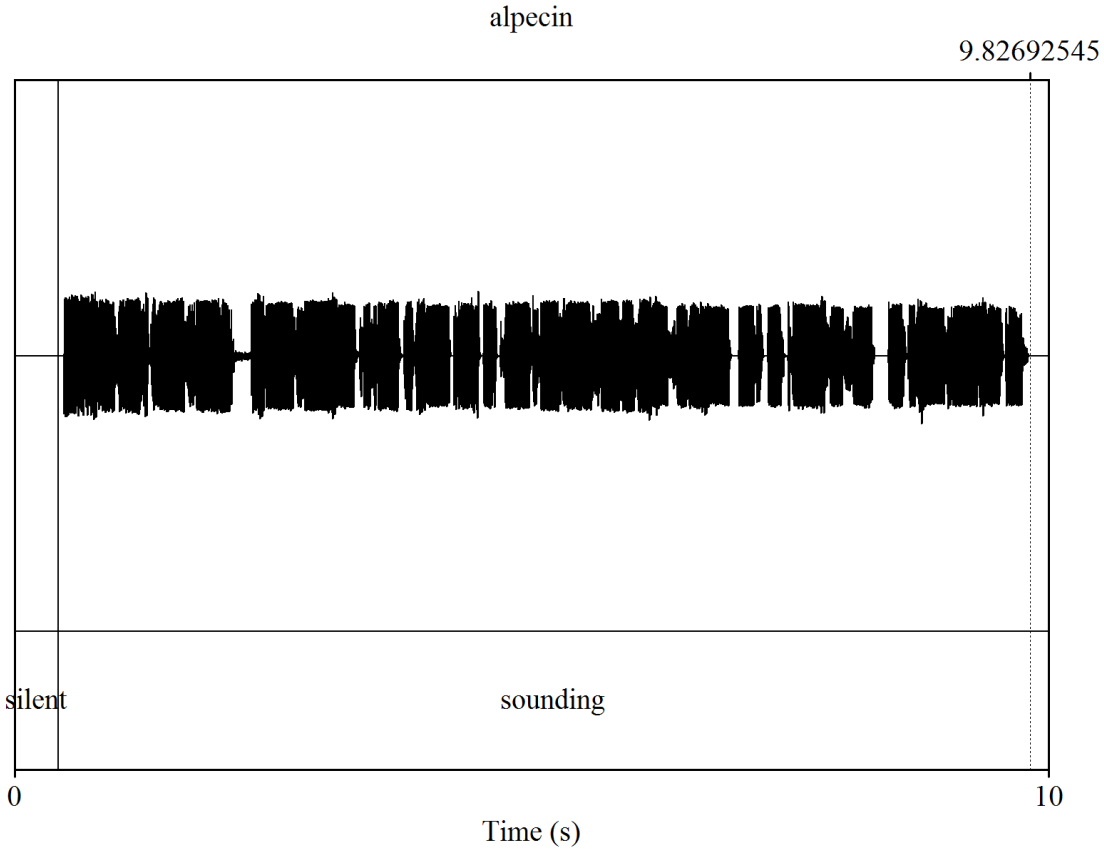
Dauer der Aufnahme: 10 sec, Sprechdauer gesamt: 9,25 sec (0,64 – 9,89)

Pausen: keine

Sprechgeschwindigkeit global: 39 Silben in 9,25 sec = 4,22 Silben /sec.

Die Hintergrundgeräusche sind bei diesem Spot eher laut. Am Anfang ist Donnern zu hören. Um die Sprache von diesen zusätzlichen Geräuschen zu trennen, musste hier der Schwellenwert von -25 dB auf -10 dB Differenz zur maximalen Lautheit verringert werden. Dadurch sind die ersten 0,64 sec mit den Hintergrundgeräuschen als „silent“ markiert. Auch dieser Spot endet akustisch bei 9,89 sec, was wiederum aufgrund der definierten Pausendauer von mindestens 0,25 sec von Praat nicht ausgewiesen ist.

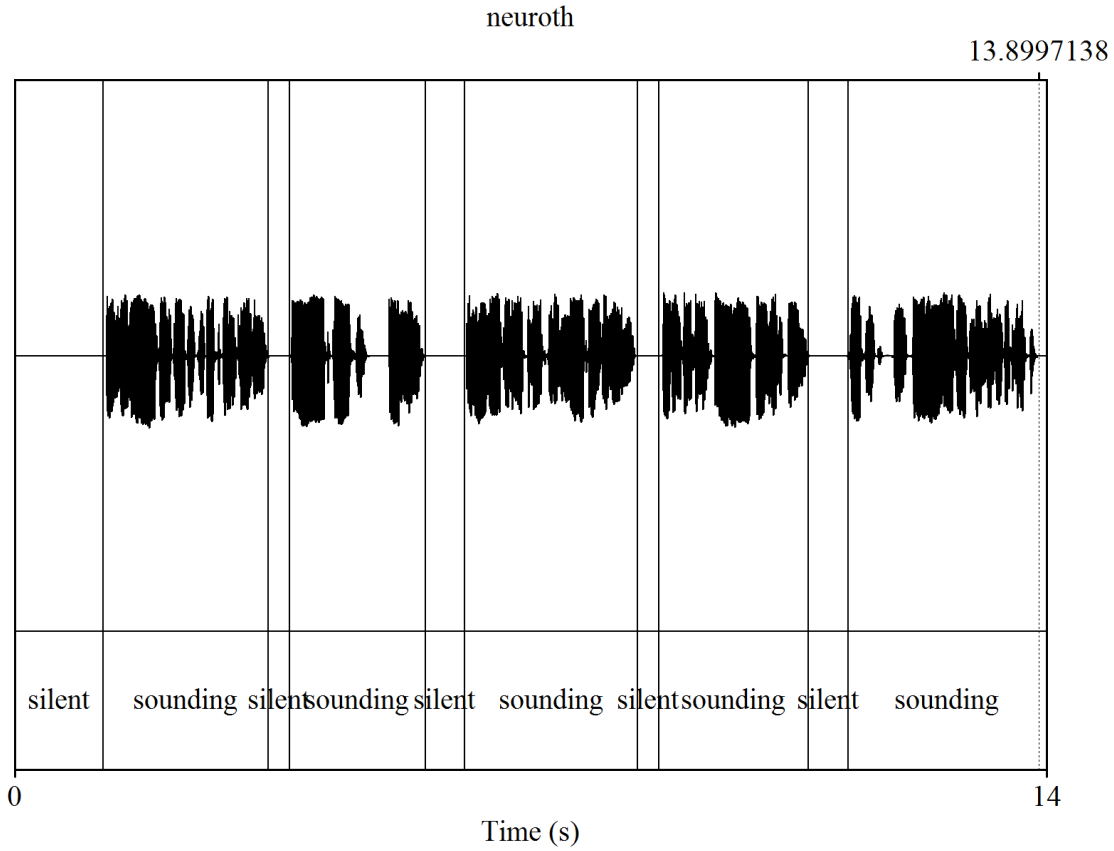
SPOT: ALPECIN



Text: „Männer färben nicht. Männer tunen. Das neue Tuning Shampoo von Alpecin kräftigt ihre natürliche Haarfarbe Wäsche für Wäsche und beugt Haarausfall vor. Alpecin Tuning Shampoo.“
Dauer der Aufnahme: 10 sec, Sprechdauer gesamt: 9,40 sec (0,43 – 9,83)
Pausen: keine
Sprechgeschwindigkeit global: 49 Silben in 9,40 sec = 5,21 Silben /sec.

Auch bei diesem Spot endet die Sprache bei 9,83 sec, ist jedoch aufgrund der definierten Pausendauer von mindestens 0,25 sec von Praat hier nicht ausgewiesen. Nach dem zweiten Satz, „Männer tunen.“, ist eine Mikropause sichtbar, die durch die Ausschwingphase des Konsonanten „n“ und den neuen Wortansatz bei „Das“ bedingt ist.

SPOT: NEUROTH



Text: „Besser hören ist jetzt Stadtgespräch. [3,44-3,73] Neuroth bringt Passion. [5,57-6,11] Kein Hörgerät trägt sich so angenehm [8,45-8,75] und ist dabei beinah unsichtbar. [10,77-11,31] Jetzt in ihrem Neuroth-Fachinstitut.“

Dauer der Aufnahme: 14 sec, Sprechdauer gesamt: 12,70 sec (1,20 – 13,90)

Pausen: 4 Pausen mit einer Gesamtlänge von 1,67 sec

Differenz Sprechdauer: 13,15%

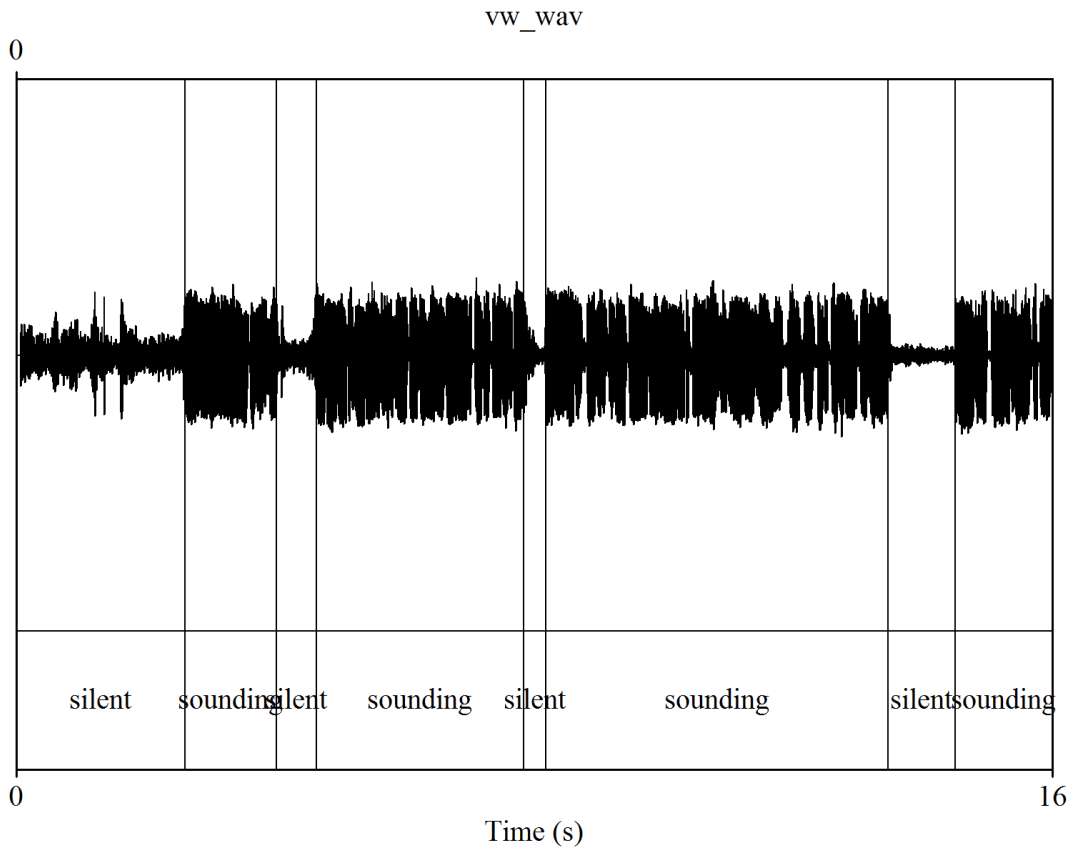
Sprechgeschwindigkeit global: 43 Silben in 12,70 sec = 3,39 Silben /sec

Sprechgeschwindigkeit minus Pausen: 43 Silben in 11,03 sec = 3,90 Silben /sec

Differenz Silben: 0,51 Silben/sec

Erwähnenswert ist bei diesem Spot die rhythmische Pausensetzung von 0,29 / 0,54 / 0,30 / 0,54 Sekunden, die bei der tontechnischen Bearbeitung der Sprachaufnahmen gezielt gesetzt werden.

SPOT: VW



Text: „Weil neu sauberer ist. [4,03-4,64] Die umweltfreundlichen VW TDI-Modelle mit Partikelfilter. [7,84-8,19] Jetzt bis zu zweitausend Euro Umweltbonus für Polo, Golf, Jetta und Passat TDI. [13,47-14,51] Näheres bei ihrem VW-Betrieb.“

Dauer der Aufnahme: 16 sec, Sprechdauer gesamt: 13,39 sec (2,61 – 16,00)

Pausen: 3 Pausen mit einer Gesamtlänge von 2,00 sec

Differenz Sprechdauer: 14,94%

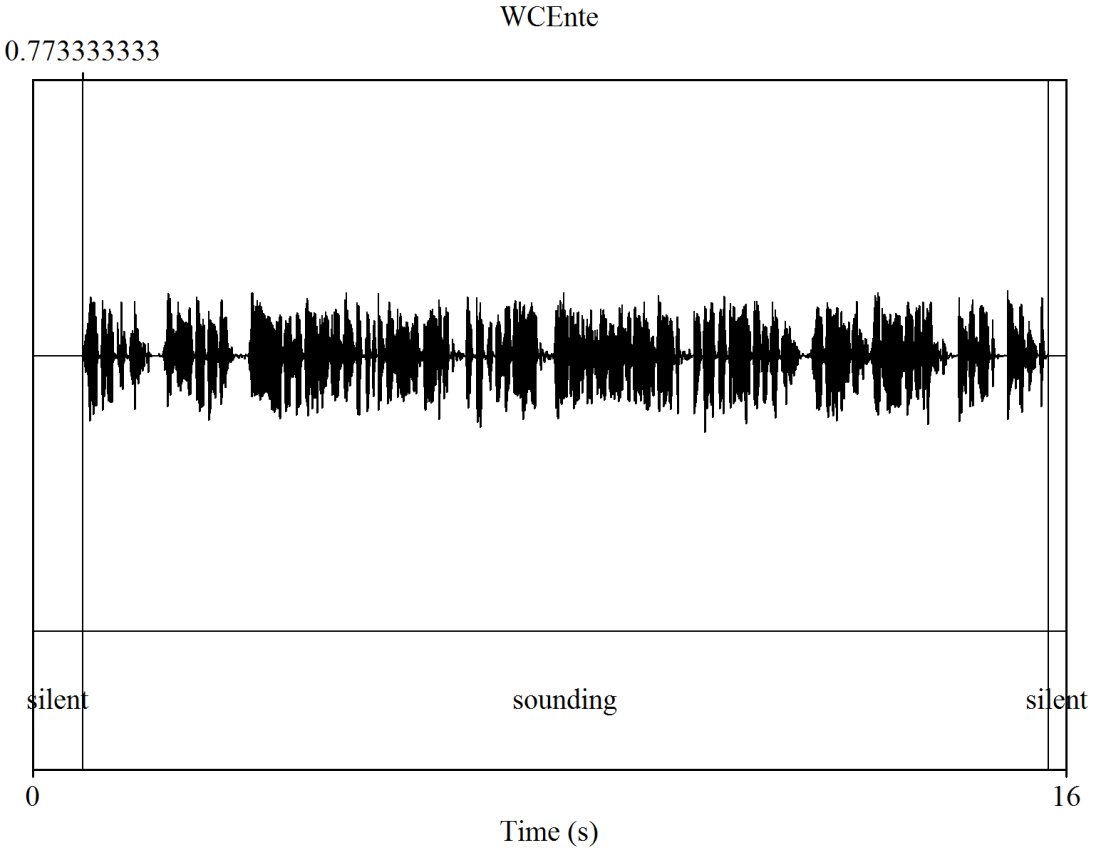
Sprechgeschwindigkeit global: 60 Silben in 13,39 sec = 4,48 Silben /sec

Sprechgeschwindigkeit minus Pausen: 60 Silben in 11,39 sec = 5,27 Silben /sec

Differenz Silben: 0,79 Silben/sec

Bei den Hintergrundgeräuschen führt insbesondere das Vogelzwitschern zu starken Ausschlägen der Amplitude. Um die Sprache davon zu trennen, musste hier der Schwellenwert von -25 dB sogar auf -5 dB Differenz zur maximalen Lautheit verringert werden. Dieses als silent markierte Vogelzwitschern ist vor allem in den ersten vier Sekunden deutlich sichtbar.

SPOT: WC-ENTE



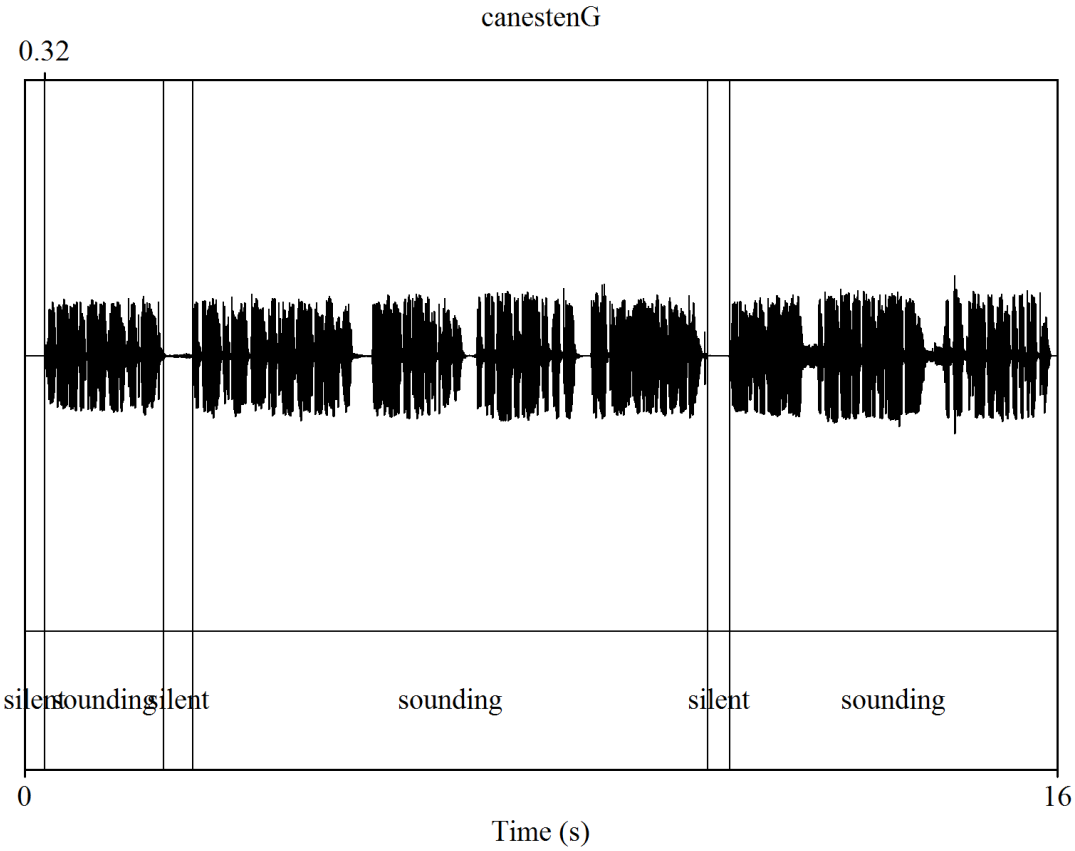
Text: „WC Putzen. Warum mit der Ente? Weil nur die WC-Ente den doppelt gebogenen Entenhals hat, der kopfüber dosiert senkrecht nach oben unter den Rand zielt und das bis zum letzten Tropfen. WC-Ente - schneller gegen Kalk. Ente gut, alles gut.“

Dauer der Aufnahme: 16 sec, Sprechdauer gesamt: 14,96 sec (0,77 – 15,73)

Pausen: keine

Sprechgeschwindigkeit global: 67 Silben in 14,96 sec = 4,48 Silben /sec

SPOT: CANESTEN GLUTRIMAZOL



Text: „Weil bei Scheidenpilz jeder Tag zählt, [2,16-2,61] gibt es jetzt Canesten Glutrimazol Gyn Once. Die neue Eintageskombi. Mit nur einer Vaginaltablette plus Creme zur äußeren Anwendung. [10,59-10,93] Schnell wieder wohlfühlen mit der Eintageskombi von Canesten. Rezeptfrei in ihrer Apotheke.“

Dauer der Aufnahme: 16 sec, Sprechdauer gesamt: 15,68 sec (0,32 – 16,00)

Pausen: 2 Pausen mit einer Gesamtlänge von 0,79 sec

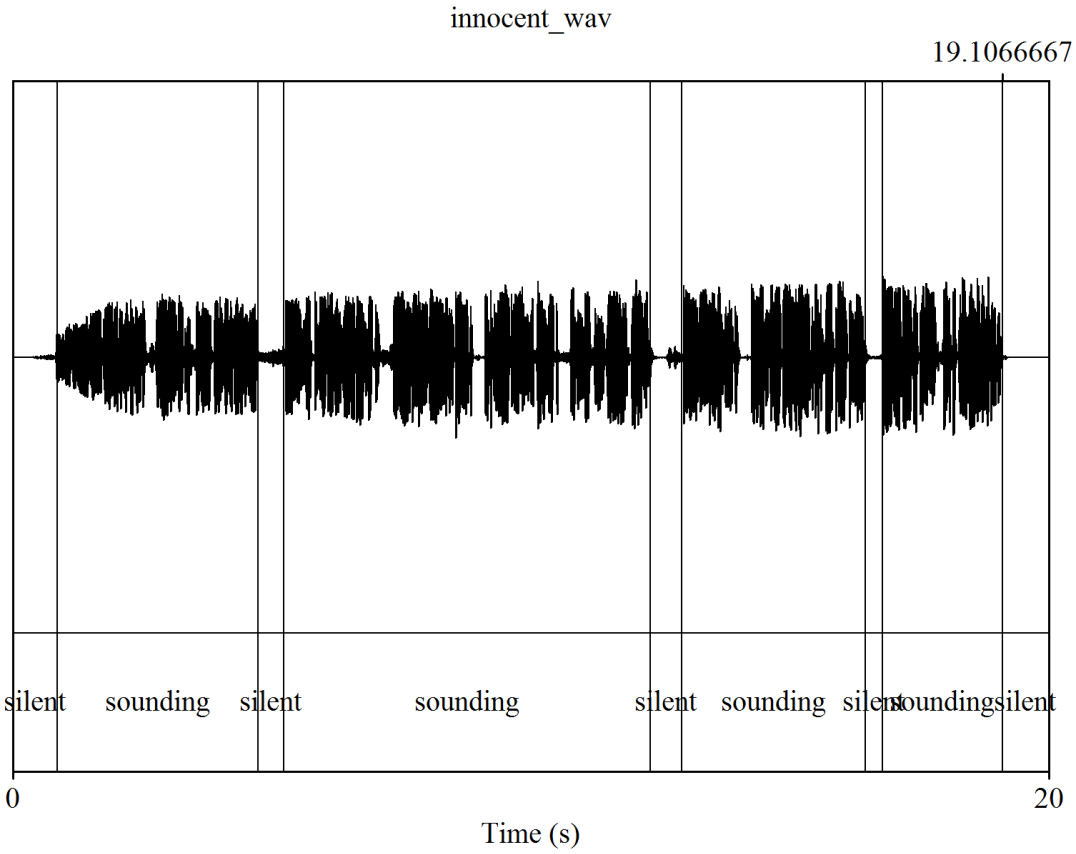
Differenz Sprechdauer: 5,04%

Sprechgeschwindigkeit global: 75 Silben in 15,68 sec = 4,78 Silben /sec

Sprechgeschwindigkeit minus Pausen: 75 Silben in 14,89 sec = 5,04 Silben /sec

Differenz Silben: 0,26 Silben/sec

SPOT: INNOCENT



Text: „Für den neuen Innocent-Saft wählen wir die besten Früchte aus. [4,73-5,24] Nur wer täglich sonnenbadet, immer an der frischen Luft ist und regelmäßig duscht, darf mit in die Saftpresse. [12,31-12,92] Innocent-Saft. Der Saft von Innocent und Mutter Natur. [16,47-16,79] Auch als Apfel- und Apfel-Himbeersaft.“

Dauer der Aufnahme: 20 sec, Sprechdauer gesamt: 18,24 sec (0,87 – 19,11)

Pausen: 3 Pausen mit einer Gesamtlänge von 1,44 sec

Differenz Sprechdauer: 7,89%

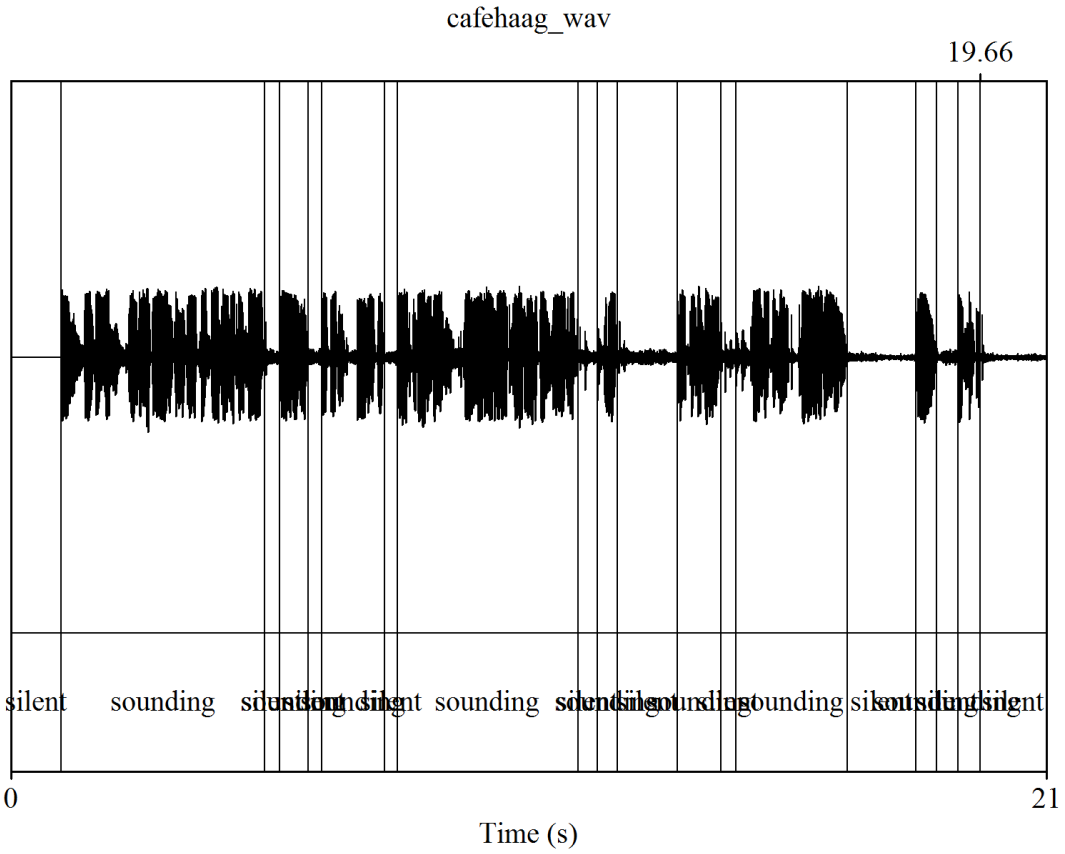
Sprechgeschwindigkeit global: 71 Silben in 18,24 sec = 3,89 Silben /sec

Sprechgeschwindigkeit minus Pausen: 71 Silben in 16,80 sec = 4,23 Silben /sec

Differenz Silben: 0,34 Silben/sec

Auch bei diesem Spot führt das Vogelzwitschern zu starken Ausschlägen der Amplitude. Um die Sprache davon zu trennen, musste der Schwellenwert von -25 dB auf -15 dB Differenz zur maximalen Lautheit verringert werden. Dieses als silent markierte Vogelzwitschern ist vor allem in der zweiten Pause deutlich sichtbar.

SPOT: CAFE HAAG



Text: „Was für mich wahre Entspannung ist, das ist einfach. Eine Tasse Cafe Haag. [5,15-5,45] Morgens, [6,03-6,30] Mittags, Abends [7,58-7,85] und auch zwischendurch. Gerade wenn ich so richtig im Streß bin, sage ich oft: [11,50-11,90] Stopp. [12,30-13,53] Zeit für Cafe Haag. [14,41-14,70] Denn Cafe Haag kann ich immer genießen. [16,97-18,35] Mmhh.[18,78-19,21]Perfekt.“

Dauer der Aufnahme: 21 sec, Sprechdauer gesamt: 18,64 sec (1,02 – 19,66)

Pausen: 8 Pausen mit einer Gesamtlänge von 4,57 sec

Differenz Sprechdauer: 24,52%

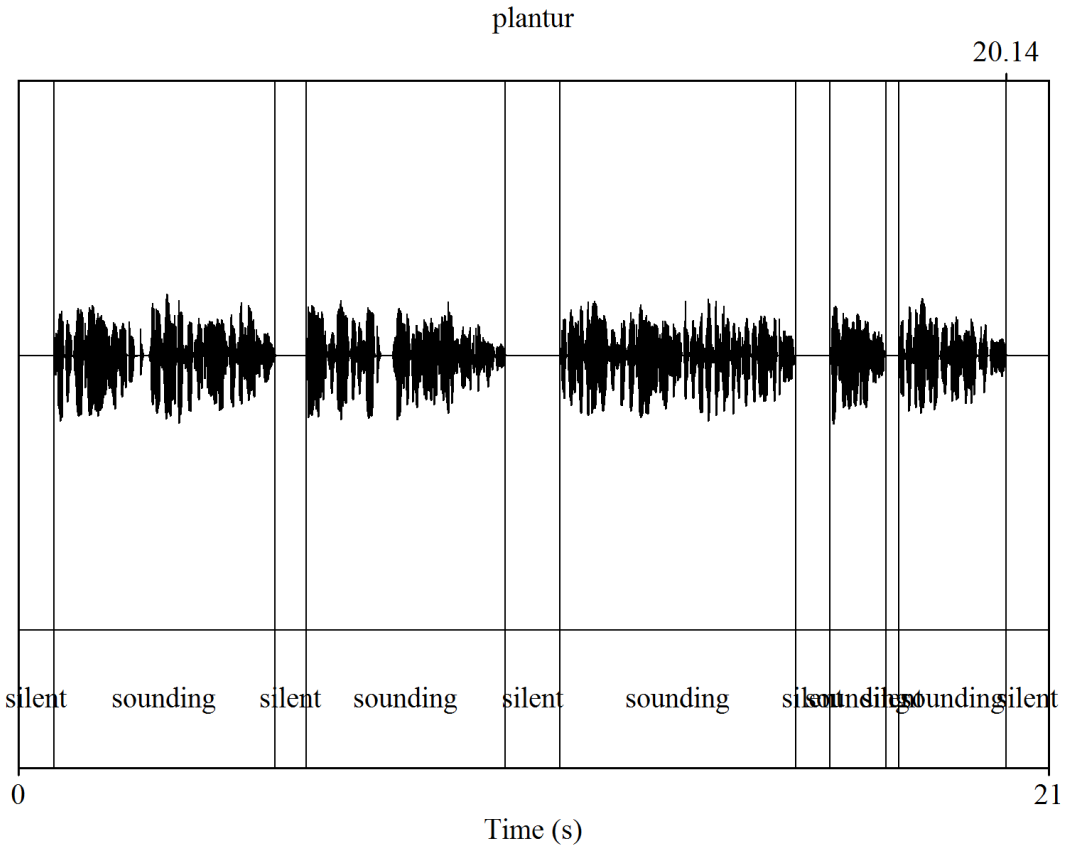
Sprechgeschwindigkeit global: 66 Silben in 18,64 sec = 3,54 Silben /sec

Sprechgeschwindigkeit mit Pausen: 66 Silben in 14,07 sec = 4,69 Silben /sec

Differenz Silben: 1,15 Silben/sec

Bei diesem Spot sind die Hintergrundgeräusche, insbesondere das Klappern von Kaffeetassen und Löffeln, sehr markant. Die Amplitudenausschläge mancher Zisch- und Plosivlaute wie zum Beispiel ein „s“ oder das „p“ bei dem Wort „Stopp“ sind sehr nahe bei denen des Geschirrklapperns. Trotzdem konnte der Schwellenwert nicht zu niedrig eingestellt werden, da dies wiederum bei anderen Pausen zu unkorrekten Anzeigen geführt hätte. Ein Schwellenwert der Differenz zur maximalen Lautheit von -17dB ergab die besten, das heißt, am ehesten korrekten Ergebnisse. Hier zeigte sich die Notwendigkeit einer auditiven Kontrolle des Rechenprogramms besonders deutlich.

SPOT: PLANTUR



Text: „Seit es Plantur neununddreißig gibt, haben Frauen über vierzig keine Angst mehr vor Haarausfall. [5,23-5,87] Wenn der Östrogenspiegel sinkt, sind schütteres Haar und Haarausfall oft die Folge. [9,93-11,05] Mit dem Koffeinshampoo von Plantur neununddreißig ist es verblüffend einfach etwas dagegen zu tun. [15,85-16,54] Plantur neununddreißig, [17,69-17,95] die Koffeintherapie für das Haar ab vierzig.“

Dauer der Aufnahme: 21 sec, Sprechdauer gesamt: 19,41 sec (0,73 – 20,14)

Pausen: 4 Pausen mit einer Gesamtlänge von 2,71 sec

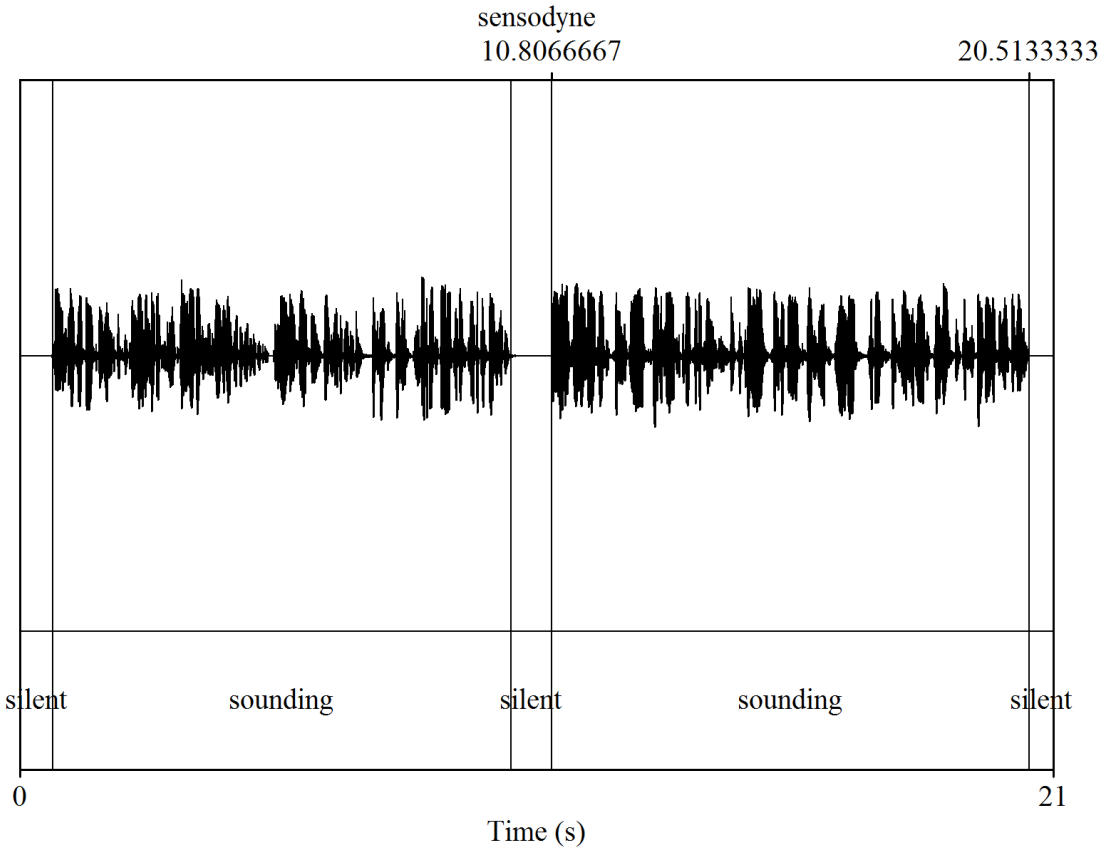
Differenz Sprechdauer: 13,96%

Sprechgeschwindigkeit global: 93 Silben in 19,41 sec = 4,79 Silben /sec.

Sprechgeschwindigkeit minus Pausen: 93 Silben in 16,70 sec = 5,57 Silben /sec

Differenz Silben: 0,78 Silben/sec

SPOT: SENSODYNE



Text: „Zahnschmelzabbau kann jeden treffen. Kinder und Erwachsene. Eine Ursache sind säurehaltige Lebensmittel. Die Säuren greifen den Zahnschmelz an. Es wird dünner. Das wichtigste ist die rechtzeitige Vorsorge. [9,98-10,81] Meine Empfehlung lautet Sensodyne Pro Schmelz tägliche Zahncreme. Und für Kinder Sensodyne Pro Schmelz Junior. Sie hilft den Zahnschmelz zu härten. Das ist Vorsorge von Anfang an.“

Dauer der Aufnahme: 21 sec, Sprechdauer gesamt: 19,84 sec (0,67 – 20,51)

Pausen: 1 Pause mit einer Gesamtlänge von 0,83 sec

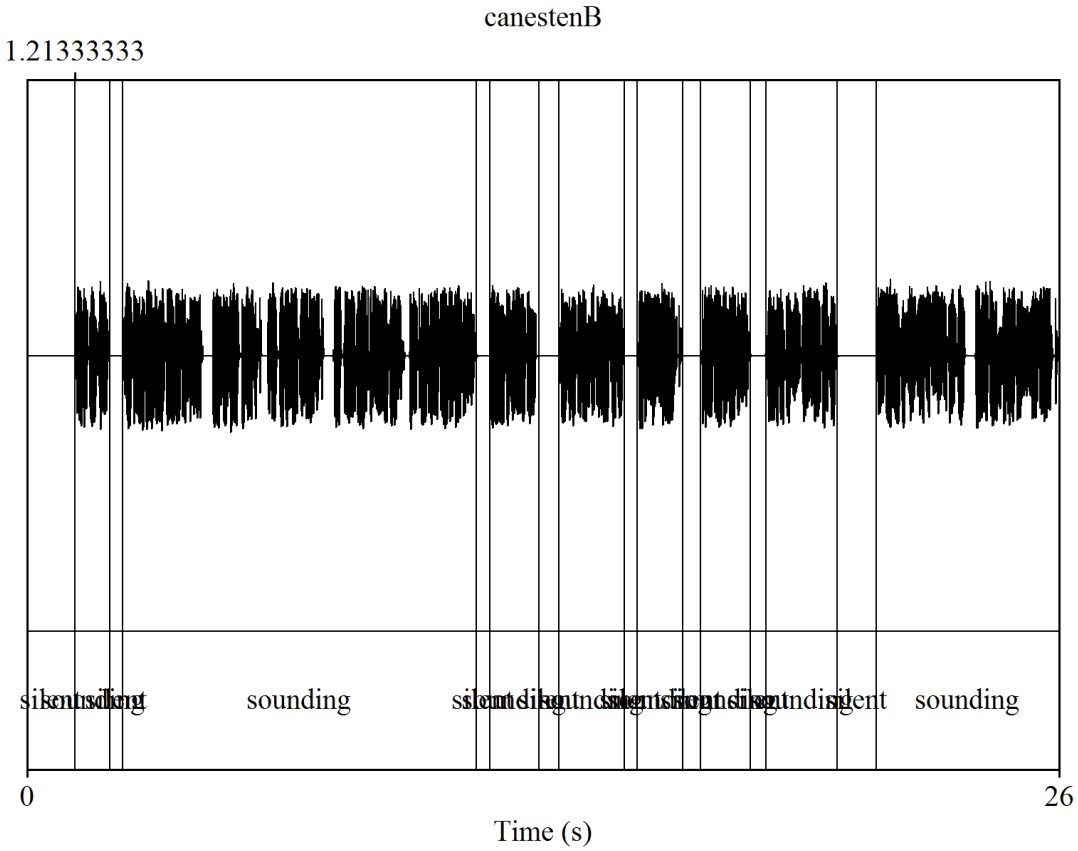
Differenz Sprechdauer: 4,18%

Sprechgeschwindigkeit global: 106 Silben in 19,84 sec = 5,34 Silben /sec

Sprechgeschwindigkeit minus Pausen: 106 Silben in 19,01 sec = 5,58 Silben /sec

Differenz Silben: 0,24 Silben/sec

SPOT: CANESTEN BIFONAZOL



Text: „Fußpilz? [2,09-2,41] Wirksame Behandlung ist besonders gründlich. Erstens, das Jucken muß weg. Zweitens, der Fußpilz muß weg. Drittens, die Haut muß sich erholen. Das kann Canesten Bifonazol. [11,32-11,67] Schnell gegen das Jucken, [12,89-13,40] hochwirksam gegen Fußpilz, [15,05-15,37] Erholung für die Haut. [16,52-16,97] Canesten Bifonazol. [18,23-18,63] Hochwirksam gegen Fußpilz. [20,41-21,40] Und schnell und gründlich gegen Nagelpilz - Canesten Bifonazol Nagelpilzset.“

Dauer der Aufnahme: 26 sec, Sprechdauer gesamt: 24,79 sec (1,21 – 26,00)

Pausen: 7 Pausen mit einer Gesamtlänge von 3,34 sec

Differenz Sprechdauer: 13,47%

Sprechgeschwindigkeit global: 100 Silben in 24,79 sec = 4,03 Silben /sec.

Sprechgeschwindigkeit minus Pausen: 100 Silben in 21,45 sec = 4,66 Silben /sec

Differenz Silben: 0,63 Silben/sec

ANHANG 2: MESSWERTE DER STIMMGRUNDFREQUENZEN IN DEN WERBESPOTS

Die folgenden Messwerte wurden mit dem Programm Praat generiert. Die Analysefenster sind zwischen 30 – 600 Hz eingestellt. Bei dieser Frequenzlänge werden die Stimmproben umfassend ausgewertet. Obertöne und Signale, die unter oder über diesen Frequenzen liegen, werden jedoch nicht erfasst. Das Messintervall (time step) ist auf 0,01 Sekunden eingestellt. Dies ergibt knapp 100 Messwerte pro Sekunde (number of frames). Die errechneten Werte sind in der Messeinheit Hertz (Hz), in der Verhältnistonhöhe Mel, sowie in Halbtonschritten (semitones) und in ERB angegeben. ERB (Equivalent Rectangular Bandwidth) ist ein gehörbezogener Schätzwert, durch den ein realistischeres Bild der Ohrfilter-Bandbreiten gewonnen werden kann als mit Frequenzgruppen. (Vgl. Terhardt 1998:255, 267) In der Auswertung der Rechenergebnisse wird mit Hertz und mit Halbtönen gearbeitet, die Nachkommastellen werden auf eine Position gerundet.

SPOT: MÖBELIX

Object type: Pitch

Object name: **möbelix_wav**

Date: Thu Jan 14 21:48:57 2016

Time domain:

Start time: 0 seconds

End time: 6 seconds

Total duration: 6 seconds

Time sampling:

Number of frames: 591 (266 voiced)

Time step: 0.01 seconds

First frame centred at: 0.049999999999999926 seconds

Ceiling at: 600 Hz

Estimated quantiles:

10% = 149.934838 Hz = 132.58793 Mel = 7.01202764 semitones above 100 Hz = 4.25150229 ERB

16% = 169.845971 Hz = 148.01544 Mel = 9.17072398 semitones above 100 Hz = 4.70789658 ERB

50% = 266.542129 Hz = 217.338126 Mel = 16.9723629 semitones above 100 Hz = 6.67824417 ERB

84% = 315.19363 Hz = 249.169279 Mel = 19.8748606 semitones above 100 Hz = 7.54385435 ERB

90% = 319.386542 Hz = 251.828258 Mel = 20.1036423 semitones above 100 Hz = 7.61515659 ERB

Estimated spreading:

84%-median = 48.74 Hz = 31.89 Mel = 2.908 semitones = 0.8672 ERB

median-16% = 96.88 Hz = 69.45 Mel = 7.816 semitones = 1.974 ERB

90%-10% = 169.8 Hz = 119.5 Mel = 13.12 semitones = 3.37 ERB

Minimum 126.312858 Hz = 113.705624 Mel = 4.04401806 semitones above 100 Hz = 3.68298541 ERB

Maximum 337.697291 Hz = 263.291885 Mel = 21.0687673 semitones above 100 Hz = 7.92086012 ERB

Range 211.4 Hz = 149.586261 Mel = 17.02 semitones = 4.238 ERB

Average: 248.085297 Hz = 202.998278 Mel = 15.0902017 semitones above 100 Hz = 6.25612238 ERB

Standard deviation: 63.24 Hz = 44.44 Mel = 4.883 semitones = 1.253 ERB

Mean absolute slope: 573.1 Hz/s = 404.9 Mel/s = 45.24 semitones/s = 11.45 ERB/s

Mean absolute slope without octave jumps: 37.2 semitones/s

SPOT: ALPECIN

Object type: Pitch

Object name: **alpecin**

Date: Fri Jan 08 01:15:51 2016

Time domain:

Start time: 0 seconds

End time: 10 seconds

Total duration: 10 seconds

Time sampling:

Number of frames: 991 (579 voiced)

Time step: 0.01 seconds

First frame centred at: 0.049999999999999926 seconds

Ceiling at: 600 Hz

Estimated quantiles:

10% = 92.9140559 Hz = 85.8510267 Mel = -1.2723748 semitones above 100 Hz = 2.82289715 ERB

16% = 108.590793 Hz = 99.1012615 Mel = 1.42682143 semitones above 100 Hz = 3.23534848 ERB

50% = 147.048917 Hz = 130.315522 Mel = 6.67555391 semitones above 100 Hz = 4.18367314 ERB

84% = 173.070208 Hz = 150.473425 Mel = 9.49628886 semitones above 100 Hz = 4.77996586 ERB

90% = 181.47694 Hz = 156.83108 Mel = 10.3174349 semitones above 100 Hz = 4.96557393 ERB

Estimated spreading:

84%-median = 26.04 Hz = 20.18 Mel = 2.823 semitones = 0.5968 ERB

median-16% = 38.49 Hz = 31.24 Mel = 5.253 semitones = 0.9491 ERB

90%-10% = 88.64 Hz = 71.04 Mel = 11.6 semitones = 2.145 ERB

Minimum 34.2865456 Hz = 33.2602845 Mel = -18.5314264 semitones above 100 Hz = 1.12023438 ERB

Maximum 556.179811 Hz = 384.312208 Mel = 29.7066165 semitones above 100 Hz = 10.9976229 ERB

Range 521.9 Hz = 351.051923 Mel = 48.24 semitones = 9.877 ERB

Average: 146.616735 Hz = 128.66274 Mel = 5.75196346 semitones above 100 Hz = 4.11476653 ERB

Standard deviation: 51.7 Hz = 36.88 Mel = 5.614 semitones = 1.068 ERB

Mean absolute slope: 375.2 Hz/s = 280.5 Mel/s = 41.1 semitones/s = 8.224 ERB/s

Mean absolute slope without octave jumps: 27.1 semitones/s

SPOT: NEUROTH

Object type: Pitch

Object name: **neuroth**

Date: Fri Jan 08 01:31:26 2016

Time domain:

Start time: 0 seconds

End time: 14 seconds

Total duration: 14 seconds

Time sampling:

Number of frames: 1391 (515 voiced)

Time step: 0.01 seconds

First frame centred at: 0.049999999999999926 seconds

Ceiling at: 600 Hz

Estimated quantiles:

10% = 56.8586786 Hz = 54.1077165 Mel = -9.77457028 semitones above 100 Hz = 1.80846426 ERB

16% = 59.5075648 Hz = 56.5031952 Mel = -8.98626017 semitones above 100 Hz = 1.88638519 ERB

50% = 112.177346 Hz = 102.088325 Mel = 1.98937635 semitones above 100 Hz = 3.32748727 ERB

84% = 151.449944 Hz = 133.777194 Mel = 7.18609249 semitones above 100 Hz = 4.28693784 ERB

90% = 157.439416 Hz = 138.453544 Mel = 7.85756127 semitones above 100 Hz = 4.42586039 ERB

Estimated spreading:

84%-median = 39.31 Hz = 31.72 Mel = 5.202 semitones = 0.9604 ERB

median-16% = 52.72 Hz = 45.63 Mel = 10.99 semitones = 1.443 ERB

90%-10% = 100.7 Hz = 84.43 Mel = 17.65 semitones = 2.62 ERB

Minimum 47.7834655 Hz = 45.8206718 Mel = -12.7849993 semitones above 100 Hz = 1.53708673 ERB

Maximum 586.411539 Hz = 399.141889 Mel = 30.6229619 semitones above 100 Hz = 11.3578237 ERB

Range 538.6 Hz = 353.321217 Mel = 43.41 semitones = 9.821 ERB

Average: 122.649068 Hz = 106.531685 Mel = 0.766272343 semitones above 100 Hz = 3.40652782 ERB

Standard deviation: 94 Hz = 64.04 Mel = 8.875 semitones = 1.815 ERB

Mean absolute slope: 372.6 Hz/s = 269.4 Mel/s = 42.28 semitones/s = 7.849 ERB/s

Mean absolute slope without octave jumps: 18.65 semitones/s

SPOT: WC-ENTE

Object type: Pitch

Object name: **WCEnte**

Date: Fri Jan 08 01:12:07 2016

Time domain:

Start time: 0 seconds

End time: 16 seconds

Total duration: 16 seconds

Time sampling:

Number of frames: 1591 (931 voiced)

Time step: 0.01 seconds

First frame centred at: 0.049999999999999926 seconds

Ceiling at: 600 Hz

Estimated quantiles:

10% = 94.6748878 Hz = 87.3553239 Mel = -0.947355461 semitones above 100 Hz = 2.87003502 ERB

16% = 99.3171514 Hz = 91.3016479 Mel = -0.118622539 semitones above 100 Hz = 2.9933111 ERB

50% = 115.623642 Hz = 104.943372 Mel = 2.51323702 semitones above 100 Hz = 3.41527091 ERB

84% = 170.420196 Hz = 148.454003 Mel = 9.22915582 semitones above 100 Hz = 4.72076823 ERB

90% = 218.73182 Hz = 184.15314 Mel = 13.5499573 semitones above 100 Hz = 5.75056851 ERB

Estimated spreading:

84%-median = 54.83 Hz = 43.53 Mel = 6.72 semitones = 1.306 ERB

median-16% = 16.32 Hz = 13.65 Mel = 2.633 semitones = 0.4222 ERB

90%-10% = 124.1 Hz = 96.85 Mel = 14.51 semitones = 2.882 ERB

Minimum 66.6327947 Hz = 62.8954822 Mel = -7.02834828 semitones above 100 Hz = 2.09319032 ERB

Maximum 590.729849 Hz = 401.227902 Mel = 30.7499821 semitones above 100 Hz = 11.4082257 ERB

Range 524.1 Hz = 338.33242 Mel = 37.78 semitones = 9.315 ERB

Average: 155.143476 Hz = 131.586674 Mel = 5.25277958 semitones above 100 Hz = 4.15552176 ERB

Standard deviation: 106.1 Hz = 71.11 Mel = 7.93 semitones = 1.979 ERB

Mean absolute slope: 676.7 Hz/s = 473.8 Mel/s = 60.76 semitones/s = 13.51 ERB/s

Mean absolute slope without octave jumps: 29.22 semitones/s

SPOT: CANESTEN GLUTRIMAZOL

Object type: Pitch

Object name: **canestenG**

Date: Fri Jan 08 01:40:32 2016

Time domain:

Start time: 0 seconds

End time: 16 seconds

Total duration: 16 seconds

Time sampling:

Number of frames: 1591 (936 voiced)

Time step: 0.01 seconds

First frame centred at: 0.049999999999999926 seconds

Ceiling at: 600 Hz

Estimated quantiles:

10% = 125.490864 Hz = 113.036745 Mel = 3.93098802 semitones above 100 Hz = 3.66263812 ERB

16% = 137.382631 Hz = 122.63505 Mel = 5.49839542 semitones above 100 Hz = 3.95323694 ERB

50% = 165.703613 Hz = 144.84132 Mel = 8.74326075 semitones above 100 Hz = 4.61457091 ERB

84% = 240.869731 Hz = 199.768243 Mel = 15.2190373 semitones above 100 Hz = 6.19045171 ERB

90% = 259.665247 Hz = 212.686434 Mel = 16.5198352 semitones above 100 Hz = 6.54981623 ERB

Estimated spreading:

84%-median = 75.21 Hz = 54.96 Mel = 6.479 semitones = 1.577 ERB

median-16% = 28.34 Hz = 22.22 Mel = 3.247 semitones = 0.6617 ERB

90%-10% = 134.2 Hz = 99.7 Mel = 12.6 semitones = 2.889 ERB

Minimum 58.2945199 Hz = 55.4074918 Mel = -9.34281395 semitones above 100 Hz = 1.85077259 ERB

Maximum 332.64378 Hz = 260.151882 Mel = 20.8077367 semitones above 100 Hz = 7.83739673 ERB

Range 274.3 Hz = 204.74439 Mel = 30.15 semitones = 5.987 ERB

Average: 180.447821 Hz = 154.492459 Mel = 9.33889437 semitones above 100 Hz = 4.87324183 ERB

Standard deviation: 55.5 Hz = 41.39 Mel = 5.737 semitones = 1.206 ERB

Mean absolute slope: 334.6 Hz/s = 247.5 Mel/s = 32.12 semitones/s = 7.171 ERB/s

Mean absolute slope without octave jumps: 25.01 semitones/s

SPOT: PLANTUR

Object type: Pitch

Object name: **plantur**

Date: Fri Jan 08 01:42:37 2016

Time domain:

Start time: 0 seconds

End time: 21 seconds

Total duration: 21 seconds

Time sampling:

Number of frames: 2091 (941 voiced)

Time step: 0.01 seconds

First frame centred at: 0.049999999999999926 seconds

Ceiling at: 600 Hz

Estimated quantiles:

10% = 122.679406 Hz = 110.742815 Mel = 3.53871702 semitones above 100 Hz = 3.59274575 ERB

16% = 127.495485 Hz = 114.666535 Mel = 4.20535385 semitones above 100 Hz = 3.71219083 ERB

50% = 152.936232 Hz = 134.941346 Mel = 7.35516282 semitones above 100 Hz = 4.3215835 ERB

84% = 198.546033 Hz = 169.517933 Mel = 11.8736824 semitones above 100 Hz = 5.33258425 ERB

90% = 226.541594 Hz = 189.712561 Mel = 14.1573115 semitones above 100 Hz = 5.90788494 ERB

Estimated spreading:

84%-median = 45.63 Hz = 34.59 Mel = 4.521 semitones = 1.012 ERB

median-16% = 25.45 Hz = 20.29 Mel = 3.151 semitones = 0.6097 ERB

90%-10% = 103.9 Hz = 79.01 Mel = 10.62 semitones = 2.316 ERB

Minimum 97.9003861 Hz = 90.1002746 Mel = -0.367362543 semitones above 100 Hz = 2.9558409 ERB

Maximum 323.257589 Hz = 254.271763 Mel = 20.3122109 semitones above 100 Hz = 7.68054868 ERB

Range 225.4 Hz = 164.171488 Mel = 20.68 semitones = 4.725 ERB

Average: 163.283597 Hz = 142.113067 Mel = 8.00845374 semitones above 100 Hz = 4.52081236 ERB

Standard deviation: 40.9 Hz = 30.55 Mel = 3.976 semitones = 0.8886 ERB

Mean absolute slope: 221.9 Hz/s = 167.7 Mel/s = 22.58 semitones/s = 4.907 ERB/s

Mean absolute slope without octave jumps: 19.74 semitones/s

SPOT: SENSODYNE

Object type: Pitch

Object name: **sensodyne**

Date: Fri Jan 08 01:45:08 2016

Time domain:

Start time: 0 seconds

End time: 21 seconds

Total duration: 21 seconds

Time sampling:

Number of frames: 2091 (1272 voiced)

Time step: 0.01 seconds

First frame centred at: 0.049999999999999926 seconds

Ceiling at: 600 Hz

Estimated quantiles:

10% = 172.09052 Hz = 149.727725 Mel = 9.39801149 semitones above 100 Hz = 4.75811996 ERB

16% = 177.320177 Hz = 153.696681 Mel = 9.91628051 semitones above 100 Hz = 4.87421051 ERB

50% = 209.605157 Hz = 177.58427 Mel = 12.8120906 semitones above 100 Hz = 5.56365888 ERB

84% = 271.224841 Hz = 220.483261 Mel = 17.2738718 semitones above 100 Hz = 6.76479108 ERB

90% = 283.749975 Hz = 228.808409 Mel = 18.0554432 semitones above 100 Hz = 6.99278104 ERB

Estimated spreading:

84%-median = 61.64 Hz = 42.92 Mel = 4.464 semitones = 1.202 ERB

median-16% = 32.3 Hz = 23.9 Mel = 2.897 semitones = 0.6897 ERB

90%-10% = 111.7 Hz = 79.11 Mel = 8.661 semitones = 2.236 ERB

Minimum 83.6470376 Hz = 77.865589 Mel = -3.09136377 semitones above 100 Hz = 2.5712946 ERB

Maximum 504.339449 Hz = 357.913301 Mel = 28.0127409 semitones above 100 Hz = 10.347972 ERB

Range 420.7 Hz = 280.047712 Mel = 31.1 semitones = 7.777 ERB

Average: 221.386291 Hz = 184.730492 Mel = 13.2641401 semitones above 100 Hz = 5.74916189 ERB

Standard deviation: 54.83 Hz = 37.67 Mel = 4.143 semitones = 1.056 ERB

Mean absolute slope: 317.7 Hz/s = 225.2 Mel/s = 26.07 semitones/s = 6.395 ERB/s

Mean absolute slope without octave jumps: 19.39 semitones/s

SPOT: CANESTEN BIFONAZOL

Object type: Pitch

Object name: **canestenB**

Date: Fri Jan 08 01:23:02 2016

Time domain:

Start time: 0 seconds

End time: 26 seconds

Total duration: 26 seconds

Time sampling:

Number of frames: 2591 (1077 voiced)

Time step: 0.01 seconds

First frame centred at: 0.049999999999999926 seconds

Ceiling at: 600 Hz

Estimated quantiles:

10% = 58.6818116 Hz = 55.7575568 Mel = -9.22817622 semitones above 100 Hz = 1.86215569 ERB

16% = 62.3565653 Hz = 59.0680518 Mel = -8.17663956 semitones above 100 Hz = 1.96955956 ERB

50% = 89.768318 Hz = 83.1533062 Mel = -1.86866078 semitones above 100 Hz = 2.73815845 ERB

84% = 116.519138 Hz = 105.682816 Mel = 2.64680319 semitones above 100 Hz = 3.43796176 ERB

90% = 129.771727 Hz = 116.511323 Mel = 4.51171326 semitones above 100 Hz = 3.76817645 ERB

Estimated spreading:

84%-median = 26.76 Hz = 22.54 Mel = 4.518 semitones = 0.7001 ERB

median-16% = 27.42 Hz = 24.1 Mel = 6.311 semitones = 0.769 ERB

90%-10% = 71.12 Hz = 60.78 Mel = 13.75 semitones = 1.907 ERB

Minimum 32.2217889 Hz = 31.3132476 Mel = -19.606698 semitones above 100 Hz = 1.05500067 ERB

Maximum 596.529826 Hz = 404.017261 Mel = 30.9191313 semitones above 100 Hz = 11.475521 ERB

Range 564.3 Hz = 372.704014 Mel = 50.53 semitones = 10.42 ERB

Average: 112.628185 Hz = 97.1274552 Mel = -1.21857348 semitones above 100 Hz = 3.10592586 ERB

Standard deviation: 107.9 Hz = 70.94 Mel = 8.812 semitones = 1.973 ERB

Mean absolute slope: 513.4 Hz/s = 358.6 Mel/s = 54.12 semitones/s = 10.3 ERB/s

Mean absolute slope without octave jumps: 22.04 semitones/s

ABSTRACT (DEUTSCH)

Die menschliche Stimme als Überbringerin (Medium) einer gesprochenen Botschaft beeinflusst dieselbe entscheidend. Der Klangcharakter einer Stimme und die Sprechweise haben einen entscheidenden Einfluss auf das Verständnis und die Interpretation einer gesprochenen Botschaft. Kommunikatoren, die Interesse an einer spezifischen Wirkung ihrer Botschaften haben, sollten daher den Einsatz bestimmter Stimmen und Sprechweisen sehr genau überlegen. In dieser Arbeit liegt der Fokus auf Stimmen, die in der Werbung eingesetzt werden. Zwei wesentliche strategische kommunikative Ziele der Werbung sind die Erregung von Aufmerksamkeit und die Vermittlung von Kompetenz. Es wird die Frage gestellt, ob bei Stimmen, die in Fernsehwerbespots eingesetzt werden, dementsprechende Kompetenz- und Aufmerksamkeitssignale identifiziert werden können. Dazu werden folgende relevante akustische Parameter der Stimmen berechnet und analysiert: die Sprechgeschwindigkeit und die Grundfrequenz mit ihren statistischen Kennzahlen Mittelwert, Median, Standardabweichung und Range. Bei der Analyse der Stimmen von acht ausgewählten Fernsehwerbespots werden drei Spots identifiziert, die eindeutige Signale entweder der Erregung von Aufmerksamkeit oder der Vermittlung von Kompetenz zeigen. Die anderen fünf Spots sind weniger eindeutig und zum Teil widersprüchlich. Um diese Ergebnisse zu ergänzen und weitergehende Aussagen machen zu können, wäre es notwendig diese Spots in einem nächsten Arbeitsschritt der auditiven Perzeption und Bewertung zu unterziehen.

ABSTRACT (ENGLISCH)

The human voice as transmitting medium of spoken messages has a marked influence on the comprehension of the messages. The character of a voice and the mode of speaking make a determining difference to the understanding and the interpretation of a message. Communicators, aiming at specific effects of their spoken messages, should therefore precisely consider how and which voices to use. In this study, the focus is laid on the voices meant to express the two strategic communicative aims of advertising, namely to gain attention as well as alertness, and to prove competence. The question is raised whether signals of alertness and excitement or of competence can be detected in voices used in audio-visual advertising spots. The following acoustic parameters of relevance for the expression of these signals are computed and analysed: the speaking rate and the fundamental frequency with its statistical indicators average, median, standard deviation, and range. As result of the signal analyses of the voices of eight selected advertising spots, three spots were detected that show explicit signals of either alertness or competence. The voice signals of the other five spots did not clearly match with one of the above mentioned communicative aims. To gain a better understanding of the effect and impact of these voice signals, further research in the field of acoustic perception would be needed.