



universität
wien

DIPLOMARBEIT

Titel der Diplomarbeit

Der emotionale Ausdruck der Stimme

Verfasserin

Maria Michaela Legenstein

angestrebter akademischer Grad

Magistra der Philosophie (Mag.phil.)

Wien, 2013

Studienkennzahl lt. Studienblatt:

A 316

Studienrichtung lt. Studienblatt:

Musikwissenschaft

Betreuer:

Univ.-Prof. Dr. Christoph Reuter

Einleitung.....	5
Abkürzungsverzeichnis.....	8
I Der emotionale Ausdruck der Stimme.....	9
1. Emotionen als Gegenstand wissenschaftlicher Forschung.....	9
1.1 Zur Definition des Begriffs „Emotion“.....	9
1.2 Basisemotionen.....	11
1.3 Evolutionstheoretischer Ansatz.....	12
2. Emotionale Lautäußerungen.....	14
2.1 Phylogenetische Entwicklung der Stimme.....	14
2.2 Frühe Emotionsbedingte Modulation der Stimme.....	15
2.3 Emotionale Laute als körperliche Vorgänge.....	16
2.4 Schallbilder emotionaler Laute nach Trojan (1948).....	17
3. Gegenwärtiger Stand der Forschung emotionaler Korrelate der Stimme.....	19
4. Ausgangspunkt für Pilotstudie.....	22
II Experiment.....	24
5. Der intendierte emotionale Ausdruck der Stimme und seine akustisch messbaren Merkmale.....	24
5.1 Abstract.....	24
5.2 Einleitung und Theorie.....	24
5.3 Methode.....	29
5.4 Ergebnisse Evaluierung der Perzepte – Rezipienten-Studie.....	34
5.5 Diskussion der Rezipienten Studie.....	42
5.6 Qualitative Erhebung - Auswertung.....	42
5.7 Diskussion: Qualitative Erhebung.....	45
5.8 Akustische Analyse – Zusammenfassung Ergebnisse.....	46
5.9 Diskussion: Ergebnisse der Akustischen Analyse und Vergleich mit den Studien.....	50
6. Conclusio.....	54
7. Glossar.....	56
8. Literaturverzeichnis.....	57
III Anhang.....	61
Anhang 1: Abstract zur Diplomarbeit.....	61
Abstract (deutsch).....	61
Abstract (english).....	62
Anhang 2: Akustische Analyse – Eine detaillierte Auswahl.....	63
Teil 1 Gesamtergebnis der lautpoetischen Wörter.....	63
Ergebnisse F0 mittels Praat.....	64
Grundfrequenzverlauf bei allen lautpoetischen Wörtern.....	65
Teil 2 Vokale.....	70
Teil 3: Spektralanalyse mittels Emapson.....	75
Anhang 3 Übersicht: Beispiele.....	87
Anhang 4 Curriculum Vitae.....	88
Anhang 5 Audiofile mit dem Hörversuch.....	91

Danksagung

All den großartigen Menschen, die mich bei der Entstehung dieser Arbeit unterstützt haben möchte ich an dieser Stelle meinen Dank aussprechen und einige davon explizit benennen.

Mein besonderer Dank gilt meinem Betreuer, Herrn Univ.-Prof. Dr. Christoph Reuter, der mit vielen Anregungen, Ideen und Kritik an der Umsetzung dieser Arbeit beteiligt war. Sein Enthusiasmus und seine inspirierende Freude an der Wissenschaft hatten während des gesamten Studiums einen motivierenden Einfluss auf mich.

Ebenso möchte ich allen Professorinnen und Professoren des Instituts für Musikwissenschaft in Wien meine dankende Anerkennung aussprechen, da sie durch ihr Engagement und die vielen interessanten Lehrangebote am Institut dieses Studium einzigartig für mich gemacht haben.

Ein herzliches Dankeschön gebührt auch Sebastian Kraync richten, den ich immer um Rat fragen durfte und der mir durch seine fachliche Kritik beim Korrekturlesen sehr geholfen hat. In weiterer Folge schulde ich meiner Schwester Katharina Legenstein und Herrn Dr. Thomas Mitterecker größten Dank, da sie mit Argusaugen das Lektorat dieser Arbeit vollendet haben. Frau Elisabeth Stephan danke ich für die Hilfe bei der englischen Übersetzung.

Persönliche Unterstützung erfuhr ich von meiner nächsten Umgebung, namentlich hervorzuheben ist hier besonders Oliver Irschitz. Für die Möglichkeit mich in die Ruhe am „Harold“ zum Schreiben zurück ziehen zu können, bekunde ich den allergrößten Dank an meine Familie, besonders meinen lieben Eltern Anna und Matthias.

Den vielen freiwilligen TeilnehmerInnen am Perzeptionstest, die sich die Zeit nahmen meine hundertvierundvierzig emotionalen Äußerungen anzuhören und zu bewerten, möchte ich abschließend ebenso danken wie auch den den SprecherInnen Tania Saedi, Marcos Valenzuela und Tristan Jorde. Letzterer hat durch seine langjährige Unterstützung bei jenen künstlerischen Projekten, welche mein besonderes Interesse am Thema „Emotion & Stimme“ geweckt haben, wesentlich zur Entstehung dieser Arbeit beigetragen.

Gewidmet ist diese Arbeit Christian Ide Hintze (Schule für Dichtung, Wien) sowie meinem Bruder Andreas.

„Die einzige Wahrheit auf der Erde ist unser Gefühl!“

(Gustav Mahler)

Einleitung

„C'est le ton qui fait la musique!“

(Trojan 1948, S. 41)

Die Motivation für dieses Thema entstand infolge von langjähriger Beschäftigung mit dem Klang der Stimme in eigenen künstlerischen Arbeiten im Genre der *Lautpoesie*. Als Inspirationsquelle galt insbesondere die *Schule für Dichtung Wien* (vgl. Schule für Dichtung) und die Begegnung mit Koryphäen auf dem Gebiet der avantgardistischen Stimmkünste. Allen voran sei hier Sainkho Namtchylak zu nennen. Peters (2004, S. 45) bezeichnet sie als eine der führenden Vertreterinnen der „extended voice“. Frau Namtchylak ist eine Obertonsängerin aus der Region Tuva (Sibirien), die in Wien wohnt und ihre Technik hier auch unterrichtet. Weiters lud die *Schule für Dichtung* Henri Chopin, einen Vertreter der *poésie sonore* -Bewegung (siehe Georgeva 2009, S. 57f. u. S. 61) mehrmals nach Wien ein. Dieses Pariser Klangkunst-Kollektiv determinierte im Zuge des Aufkommens des Tonbands Ende der 1950er Jahre die Schriftkultur als obsolet. Henri Chopin arbeitete nicht nur mit seiner Stimme, sondern baute durch besondere Mikrofonierung die Geräusche des Körpers in seine Performances und Stücke ein. „Je n'ai jamais accepté la poésie sans voix et sans corps!“ („Ich habe niemals Poesie ohne Stimme und ohne Körper akzeptiert“) (Zitat Henri Chopin in: de Vree 1968, S. 9).

Eine andere Art sich auf den akustischen Nucleus von Sprache zu konzentrieren und daraus Poesie zu kreieren ist bei Prof. Gerhard Rühm (Wiener Gruppe) zu finden. Ein radikaler *konkreter* Poet (Georgeva 2009, S. 58ff.) dessen akribisch akademische Auseinandersetzung mit Kunst und Sprache in einem großartigen Oeuvre an Publikationen resultiert (vgl. Schule für Dichtung/Gerhard Rühm). Der ehemalige Leiter der *Schule für Dichtung Wien* Christian Ide Hintze brachte in seinen akustischen Gedichten durch Reduktion der Sprache Poesie auf den Punkt (siehe Hintze). Er hatte auch die Idee, für diese unzähligen Genres und Bezeichnungen den Sammelbegriff *Akustische Poesie*, der auch schon von anderen genannt worden ist, zu verwenden. In den Jahren 2009/10 initiierte er die Klasse „Akustische Poesie“ an der *Schule für Dichtung* (siehe: Akustische Poesie), bei der die StudentInnen sogar mit den kleinsten Partikeln der Sprache lernt, den

Phonemen lernten, Kompositionen zu erstellen.

Es kursieren natürlich unterschiedliche Begriffe für die verschiedenen Strömungen, Techniken, Sub-Genres und Sezessionsbewegungen der akustischen, poetischen Kunst. Oft handelt es sich bei einem Begriff einfach um ein Label, das sich ein Künstler/eine Künstlerin selbst verliehen hat oder um die Übersetzung aus einer anderen Sprache. Die Begriffs-Debatte würde den Rahmen dieser Arbeit sprengen. Weiterführendes zum Thema *Lautpoesie* findet sich bei Lentz (2000).

„Innerhalb der sich konturierenden Begriffsgeschichte, in der *expressis verbis* oder zumindest tendenziell ‚die‘ Lautpoesie als impulsgebender Vorläufer der ‚Akustischen Kunst‘ zugerechnet wird [...], kann grundsätzlich unterschieden werden zwischen individuellen Begriffsgebungen (Neologismen: u. a. Wolman, Dufrêne), die ausschließlich mit dem Werk des einzelnen Autors bzw. einer einzelnen Autorin in Verbindung zu bringen sind – darüber hinaus sich aber durchaus in überindividuelle Zusammenhänge reintegrieren ließen –, und solchen Bezeichnungen, die, obwohl sie zunächst vielleicht auch ‚nur‘ in einem individuellen Kontext Verwendung fanden, sich als z. T. auch international akzeptierte Genre- oder Gattungsbezeichnungen bzw. als verschiedene Strömungen innerhalb des jeweiligen Genres kennzeichnende Differenzbegriffe durchsetzen konnten, wie z. B. *text-soundcomposition* oder – im Zusammenhang mit Chopin – *poésie sonore*.“
(Lentz 2000, S. 50f.)

Allen gemeinsam ist der besondere Fokus auf den Klang der Stimme und die Intention vor allem durch akustische Mittel Sinn zu stiften und seien es oft nur Gefühlsregungen, die evoziert werden. Die musikalischen Qualitäten (Intonation, Rhythmus, u.a.) von Sprache stehen somit im Vordergrund. Semantik hat bei den meisten Richtungen keine, bei manchen nur sehr wenig Bedeutung. Dichtkunst bezeichnet Georgeva (2009, S. 67) in ihrer Arbeit über die *Schule für Dichtung* als „Sprache der Emotionen“.

Da der klangliche Eindruck oft mehr über die wahre Intention des Sprechers/der Sprecherin enthüllt, als Sprache es aussagen zu vermag beinhaltet Gesangs- und Sprachtraining nicht selten physiologische wie auch psychische Komponenten (siehe Trojan (1955 und 1948). Ebenso informieren Erfahrungsberichte aus der Praxis von SchülernInnen der Stimmkunst und deren LehrerInnen, wie z.B. der amerikanischen Gesangslehrerin und Sängerin Dolli Melaine (Melaine 2013) darüber. Diese Annahme findet Evidenz in den vielzähligen Arbeiten zur psychokatharsischen Wirkung von Stimmarbeit und der Bedeutung von Stimmtraining für die Persönlichkeit sowie zur Vorbeugung von Stimmkrankheiten (siehe z.B. Quint 2005). Da der psychologische wie auch gesundheitliche Aspekt von Stimmtraining keine Schwerpunkte dieser Arbeit darstellen, wird in weiterer Folge auch nicht darauf

eingegangen. Es sei aber angemerkt, dass die (vor allem positiven) Wirkungen auf den Menschen bei dieser Diplomarbeit immer mitschwingen.

Die Resonanz auf die eigenen künstlerischen Arbeiten, selbstständig oder als Lautpoesie-Duo *LegéNi* (siehe: LegéNi) zeugt großteils von einem Verständnis für diese besondere Art von Kunst und ihrer suprasegmentalen (siehe Glossar, Kap. 7 S. 56) Kommunikation. Dass Stimmkunst in Form der gängigen Genres wie Klassik, Rock, Pop, etc. emotional berührt, davon zeugen unzählige Werke (siehe z.B. Siegwart/Scherer 1995, u.v.a.), doch es scheint, als könnte man die Menschen auch mit dieser – sehr archaisch anmutenden – Kunst Lautpoesie berühren können.

Von den oben genannten Einzelfällen (künstlerische Aktivitäten) sollte diese Diplomarbeit zu einer Sammlung von Tatsachenwissen werden. Der Hauptteil dieser Diplomarbeit besteht aus einem explorativen Experiment (siehe Kap. 5 und Anhang 2) mit Hörversuch und anschließenden akustischen Analysen. Im Vorfeld wurde nach Studien zur Detektion von Emotionen in der Stimme gesucht und im Bereich der Spracherkennung fündig (siehe Kap. 3).

Um bei meinen künstlerischen Aktivitäten optimale Sprachaufnahmen zu erzielen, beschäftige ich mich seit mehreren Jahren sehr genau mit stimmlichem Ausdruck, sei es in praktischer wie auch theoretischer Form. Somit habe ich großes Interesse daran zu untersuchen, ob die/der gelernte SprecherIn Emotion übermitteln kann und der Hörversuch zeigt, welche Äußerungen auch dieselben Emotionen bei den Rezipienten evozieren. Die elektroakustische Analyse als ein wichtiges Verfahren zur auditiven Beurteilung der Stimme soll in weiterer Folge eine objektive Erfassung von stimmlichen Parametern erzielen. Die eigenen Erkenntnisse werden dann mit jenen aus den Studien von Banse/Scherrer 1996 und Yildirim *et. al.* 2004 verglichen. Zwar fließen natürlich andere Komponenten (z.B. psychologische) in diese Arbeit mit ein, doch wird die Rezeption im Detail nicht untersucht. Allem voran steht die schwierige Frage: Wie kann der emotionale Ausdruck in der Stimme gemessen werden?

Es stellte sich heraus, dass die Thematik „Emotion in der Stimme“ wesentlich komplexer und schwieriger zu erfassen war, als im Vorhinein angenommen. Die Fülle an Forschungsansätzen ist dabei so groß, dass eine Fokussierung anfangs unmöglich erschien, was jedoch bei so einem weitreichenden Themengebiet jedoch unerlässlich ist. Im Laufe der Arbeit gelang es dann diese Eingrenzung zu bewerkstelligen.

Die vorliegende Diplomarbeit soll als eine wissenschaftliche Grundlage zur Optimierung von Kunst-Installationen mit Schwerpunkt Emotion in der Stimme dienen, Zielsetzung hierbei ist die Schaffung interaktiver Räume mit emotionalen Phonemen. Diese Räume sollen eine Chance zur Besinnung auf die eigene emotionale Wahrnehmung und die Wirkung auf andere sein. Das Experiment (siehe Kap. 5) bietet die Möglichkeit einzelne Sprachsounds in Isolation und nicht in einer Konversation zu analysieren.

Als Vorbereitung wurden Studien zur Erkennung von Emotion in der Sprache gesucht. Die Vielzahl an Studien in diesem Themengebiet (siehe Kap. 3 und 5.2) zeugen von großem kommerziellen Interesse seitens der Industrie. Weiters liegt eine Fülle an Studien zur Emotionsdetektion in Sprache z.B. für die Spionage vor, welche öffentlich nicht zugänglich sind. Erkenntnisse der Wissenschaft werden leider zu oft auch für den militärisch-geheimdienstlichen Sektor missbraucht. Joseph Weizenbaum, ein IT-Pionier, der 1966 das Sprachanalyseprogramm ELIZA entwickelte, wurde zum „Häretiker seiner Zunft“ (oe1 2013), da „zu viele IT-Programme [...] im Dienst militärischer Interessen [standen]“ (oe1 2013).

Neben den oben genannten Bereichen Spracherkennung und Kunst sind Studien über Emotion in der Stimme noch in vielen anderen Anwendungsgebieten von großem Nutzen – um hier einige zu nennen: Phonetik, Stimmanalyse und –aufnahme, Rhetorik, Systematische Musikwissenschaft (Mensch/Wirkung, Musikpsychologie und –soziologie). Das Wissen um den emotionalen Gehalt von Sprache ist – und das darf vor allem in der Alltagsanalyse und Reflexion nicht vergessen werden – für die Kreation von Phantasiewörtern in der Werbung bedeutend.

Die vorliegende Arbeit fokussiert jedoch im speziellen auf die soziale Funktion von Emotion die im alltäglichen Leben wie auch in der Sprache nicht wegzudenken ist.

Abkürzungsverzeichnis

Abb.	Abbildung
IPA	Internationales Phonetisches Alphabet
Kap.	Kapitel
Spr.	Sprecherin, Sprecher, SprecherInnen
Vp	Versuchsperson, Mehrzahl: Vpn
Pb	Proband
VI	Versuchsleiterin

I Der emotionale Ausdruck der Stimme

1. Emotionen als Gegenstand wissenschaftlicher Forschung

1.1 Zur Definition des Begriffs „Emotion“

Emotionen werden durch Reize der Umwelt ausgelöst und das menschliche Gehirn assoziiert diese mit bestimmten Erfahrungen.

„*Emotionen* sind erlebniszentrierte Antworten des Organismus, die die Relevanz eines Erkenntnisgegenstandes für die Befriedigung von Bedürfnissen widerspiegeln (z. B. nach den Kriterien „förderlich“ oder „hinderlich“) und verschiedene kognitive und motivationale Systeme im Sinne einer optimalen Bedürfnisbefriedigung aktivieren oder hemmen.“

(Kuhl 2010, S. 22)

Der emotionale Zustand wird memoriert und diese Erinnerung wird bei einer wiederholten Reizung abgerufen, auf die wiederum eine entsprechende Verhaltensantwort folgt. Die Emotionen beschreiben ein Kontinuum an mentalen Zuständen.

Bei Juslin/Sloboda 2010 wird der Begriff *Emotion* noch detaillierter als kurze aber sehr intensive affektive Reaktion (vgl. Juslin/Sloboda 2010, S. 10), die für gewöhnlich noch mehrere durchwegs synchronisierte Teilelemente in sich birgt: „[...] subjective feeling, physiological arousal, expression, action tendency, and regulation [...]“ (Juslin/Sloboda 2010, S. 10) definiert. Dagegen fungiert *Affekt* als Überbegriff für „all evaluative—or ‚valenced‘ (positive/negative)—states (e.g. emotion, mood, preference).“ (Juslin/Sloboda 2010, S. 10) – ein Terminus der das Phänomen im Allgemeinen beschreibt (vgl. Juslin/Sloboda 2010, S. 10). Bei Darwin (1874) wiederum dient „Emotion“ als generelle Definition.

Bei der Definitionsfrage sind sich nicht alle Wissenschaftler einig und sie erweist sich als durchaus komplex (siehe Kleinginna 1981). Doch scheint sich heutzutage *Emotion* als Sammelbegriff etabliert zu haben (vgl. Kleinginna 1981, S. 355).

„Gefühle (Emotionen) sind Reaktionsmuster auf 3 Verhaltensebenen (subjektiv, physiologisch, motorisch), die Annäherung oder Vermeidung auslösen und mit unterschiedlicher Erregung einhergehen.“ (Birbaumer/Schmidt 2010, S. 712).

Emotionen haben feststellbare Effekte auf den Körper sowie auf das Verhalten und erfüllen lebenswichtige Funktionen. Sie lösen Energetisierungsprozesse aus, die beim Menschen rudimentäre Verhaltensmuster aufbrechen und durch neuere

ersetzen. Emotionen kann man daher als Ausdrucksformen intuitiver Verhaltensmechanismen (Flucht bei Furcht, Ekel und Abwehrreaktion) betrachten (siehe Kuhl 2010, S. 103): „Ohne die Beteiligung intuitiver Verhaltensprogramme wäre weder diese verhaltenssteuernde noch die den Emotionsausdruck modulierende Wirkung von Emotionen vorstellbar [...].“ (Kuhl 2010, S. 103) Bei Emotionen kann man einen automatischen Mechanismus beobachten, da die Reaktionszeit auf einen Stimulus oft sehr kurz ist:

„Emotions can begin so quickly that they can happen before one is aware that they have begun. Quick onset is central to the adaptive value of emotions, mobilizing us quickly to respond to important events.“

(Ekman/Davidson 1994, S. 16)

Emotionen sind nach ihrer positiven und negativen Valenz unterscheidbar (vgl. Kuhl, S. 108).

Bei der Betrachtung von Emotionen stößt man unweigerlich auf die Begriffe *Kognition* und *Motivation*, da diese gemeinsam mit *Emotion* auf verschiedenen Ebenen zusammenwirken:

„Mit dem Begriff *Kognition* werden alle Prozesse zusammengefasst, die dem ‚Erkennen‘ dienen, also der Aufnahme und Verarbeitung von Wissen über die Außen- und Innenwelt (z. B. Wahrnehmung, Aufmerksamkeit, Gedächtnis und Denken). [...]. Der Begriff *Motivation* fasst Prozesse zusammen, welche an der Vorbereitung und Durchführung von Handlungen beteiligt sind, die Bedürfnisse befriedigen oder ihre Frustration vermeiden sollen.“

(Kuhl 2010, S. 22)

Von den Emotionen zu unterscheiden sind die *Triebe*, die biologische Grundbedürfnisse in z. B. Ernährung (Hunger, Durst etc.) oder Sexualität homöostatisch steuern: „Homöostase ist die Tendenz des Organismus alle lebenswichtigen Zustände einigermaßen konstant zu halten.“ (Kuhl 2010, S. 190) Weiters sind *Emotionen* von Begrifflichkeiten wie *Gewohnheiten*, *Temperament* und *Motive* zu unterscheiden (siehe Kuhl 2010, S. 28f.).

Als *Stimmungen* werden „länger anhaltende (Stunden, Tage) emotionale Reaktionstendenzen [bezeichnet], die das Auftreten einer bestimmten Emotion wahrscheinlich machen (gereizte Stimmung führt z. B. häufiger zu Ärger).“ (Birbaumer/Schmidt 2010, S. 712). Die subjektive Beschreibung von Emotionen oder Stimmungen wird bei Juslin/Sloboda 2010 als „Feeling“ (vgl. Juslin/Sloboda 2010, S. 10) benannt.

Wie wichtig die Äußerung von Emotionen ist, erkennt man oft erst, wenn diese unterdrückt werden. Während die linke Hemisphäre unseres Gehirns mit der bewussten Kontrolle beschäftigt ist und kaum mit den affekterzeugenden Systemen

zusammen hängt, ist die rechte Hälfte sehr stark mit ihnen verknüpft (Niedenthal 2007, S. 315f.). Das hat natürlich zur Folge, dass emotionsrepressive Handlungen andere Regionen durch Kompensierung in Mitleidenschaft ziehen und auf die Dauer zu keinen Lösungen führen (Frankl 2002).

1.2 Basisemotionen

Nach Kuhl (2010) können Emotionen „in [...] differenzierterer Weise in klar abgrenzbare und verbal benennbare Kategorien eingeteilt werden.“(Kuhl 2010, S. 108). In der Literatur kristallisieren sich zwei Grundsysteme von Emotionen heraus, zum einen das sozial-kognitive System (d. h. erlernt, kulturell ausdifferenziert und entwickelt wie z. B. Neid und Liebe), zum anderen jenes der Basisemotionen, wie Freude und Angst.

Ein essentielles Werk zur Emotionsforschung ist die von Ekman & Davidson herausgegebene Publikation „The nature of Emotion“ (siehe Ekman/Davidson 1994). Im ersten Kapitel geht Paul Ekman der Frage nach ob es Basisemotionen überhaupt gibt: „being basic is a property of our concepts, not of emotions per se“ (Ekman/Davidson 1994, S. 7) Diese Kategorisierung sei laut Ekman ein von den Menschen erdachtes Hilfsmittel, um durch die kategorische Einteilung von Emotionen zu Prototypen zu kommen. Diese würden wiederum helfen, Wissen über Phänomene zu generieren (vgl. Ekman/Davidson 1994, S. 9). Ekman bezeichnet Basisemotionen als „those that fulfill vital functions“ (Ekman/Davidson 1994, S. 12) und diese lebenswichtigen Funktionen können nach drei verschiedenen Kriterien – biologisch, sozial und psychologisch – bemessen werden (siehe Ekman/Davidson 1994, S. 12). Da es wichtige Emotionen – deren Bedeutung sich oft über die Zeit ändert – in anderen Kulturen gäbe, die in unserer nicht existieren, also sozial konstituiert sind (siehe Ekman/Davidson 1994, S. 12), wenn Basis Emotionen universal für alle Menschen wären, so sind sie auch „observable in rudimentary form in nonhuman primates [...]“ (Ekman/Davidson 1994, S. 12). Die Definition von Basisemotionen ändere sich auch im Laufe der Zeit: „During the Middle Ages, hope was classified as a basic emotion; today it is regarded as secondary or derived by most emotion theorists, if it is regarded at all [...]“ (Ekman/Davidson 1994, S. 12).

Die Einteilung ist im wissenschaftlichen Kontext heftig umstritten, gestaltet sich als nicht sehr einfach und doch behauptet sich die Klassifizierung von „Basisemotionen“ und deren Interkulturalität (siehe Ekman 1988) in der Literatur (siehe Reizenzein

2000). Schon bei Descartes (siehe Descartes 1996) sind primäre Affekte (Freude, Hass, Traurigkeit, Verwunderung, Begehren, Liebe) zu finden. Einige dieser Basisemotionen überdauerten die Zeit. Viele der Auflistungen der Grundemotionen enthalten „Angst, Ärger und Traurigkeit, die meisten inkludieren Freude, Liebe und Überraschung“ (Plutchik 2001, S. 349) Bei Ekman (1988) sind es Freude, Wut, Frucht, Verachtung, Traurigkeit, Überraschung und Ekel. Dornes (1995) beschrieb folgende Basisemotionen: Freude, Ärger, Traurigkeit, Furch, Interesse/Neugier, Überraschung, Ekel, Schuld und Scham. Eine weitere Zusammenfassung findet sich bei Paeschke 2003 (S. 58f.) der am häufigsten genannten Basisemotionen: „Freude, Ärger/Wut/Zorn, Angst/Furcht und Trauer“ (Paeschke 2003, S. 59).

Da sich die meisten Studien, die für die Diplomarbeit recherchiert wurden, auf das Konzept der Basisemotionen berufen, werden auch in dem vorliegenden Experiment (siehe Kap. 5) primäre Emotionen untersucht.

Doch können allgemein beschriebene Empfindungen natürlich individuell differenzieren:

„Die Selbstverständlichkeit der Emotionsklassifikation darf nicht darüber hinwegtäuschen, dass es sich ähnlich wie bei der Objekterkennung in der visuellen Wahrnehmung um einen aktiven Konstruktions-, Abstraktions- und Kontrastierungsprozess handelt: Eine mit demselben Wort benannte Emotion kann bei verschiedenen Menschen durchaus ein ganz unterschiedliches Muster physiologischer, kognitiver und verhaltensseitiger Merkmale bezeichnen.“

(Kuhl 2010, S. 110)

1.3 Evolutionstheoretischer Ansatz

Die Evolutionstheorie als wissenschaftliches Denksystem hat es zuwege gebracht, dass sich der systematisch-rationale Zugang europäischer – eurozentrischer – akademischer Forschung der kulturellen Entwicklung der – im 18./19. Jahrhundert sogenannten – „exotischen“ Völker geöffnet hat. In weiterer Folge verstand man darin Vorstufen der eigenen Entwicklung. Damit konnte der Versuch einer Rekonstruktion der Ursprünge aller menschlichen Kultur unternommen werden.

Der evolutionstheoretische Ansatz in der Tradition Darwins besagt, dass Emotionen phylogenetisch bedingt seien und das Leben sozial lebender Spezies ermögliche. Ein weiterer Vertreter der phylogenetischen Denkrichtung, Paul Ekman (siehe Ekman/Davidson 1994 u. Ekman 1988) schließt daran an und behauptet, dass uns Emotionen helfen „to deal with fundamental life tasks in ways that have been adaptive phylogenetically (Ekman/Davidson 1994, S. 16)“.

Der evolutionstheoretische Ansatz – den Ursprung von Emotionen suchend – basiert auf der Evolutionstheorie, die ihre Begründung im Hauptwerk Darwins „On the Origin of Species“ 1859 (siehe Darwin 2009) findet. Darin publiziert er seine wissenschaftliche Theorie der „natürlichen Selektion“. Dieses Buch und die darin enthaltenen Ansätze führender Wissenschaftler (Humboldt; Spencer u. v. a.) waren für die damalige Zeit revolutionär und für weitere Forschungen wegweisend.

Die phylogenetische Denkweise (Ekman 1994; Darwin 1874, Hauser 1996) bekundet, dass es universelle Emotionen und somit auch einen universellen Ausdruck von Emotionen gibt. Daraus leitet die Linguistik die Existenz universeller Basisemotionen ab.

Die Gedankengänge Darwins weiterführend, erkennt auch Trojan (1952) den universalen Charakter emotionaler Äußerungen an, doch „[...] gegen ihre exakte Übereinstimmung sprechen gewichtige Gründe. Zweifellos gibt es auf verschiedenen Kulturstufen graduelle Unterschiede in der Intensität des Gefühlsausdruck.“ (Trojan 1948, S. 39)

Weiters, so Trojan (1948),

„[...] beweisen Einzelbeobachtungen leicht nationale Unterschiede des Affektausdrucks: zornige Erregung äußert sich etwa bei vielen Norddeutschen durch erhöhte Stimmbandpressung und verschärften coup de glotte, wogegen etwa der Wiener beim gleichen Affekt die Kompression eher verringert und die Vokale mit einem rauhen Stridor hervorbringt. Unvergleichlich ist auch der Ausdruck des behaglichen Entzückens im Englischen (besonders bei Frauen) oder der der willkürlich ausgedrückten Ekstase im Französischen. Keinem Deutschen wird es einfallen, Höflichkeit und Ergebenheit mindestens in dem Maße durch Verlangsamung und Vokaldehnung auszudrücken, wie dies etwa der Serbe oder der Ungar tut. Zu bedenken ist bei solchen Vergleichen allerdings, daß auch die zugrundeliegenden Affekte wesensverschieden sein dürften.“

(Trojan 1952, S. 40)

Zu dieser Feststellung kommt auch Augustin Nebert in seinem Artikel „Tonhöhe und Sprechstimme – Unterschiede in Mutter- und Fremdsprache“ (2007). So hat die „[...] Muttersprache Einfluss auf die Sprechstimmlage [...].“ (Nebert 2007, S. 2) Es kann daher zu „[...] emotionalen Interferenzen [...]“ (Nebert 2007, S. 1) kommen, wenn man den falschen Tonfall in einer Sprache benutzt. Nebert (2007) plädiert dafür, unbedingt die mittlere Sprechstimmlage im Fremdsprachenunterricht zu lehren, da diese sprachgebunden sei (vgl. Nebert 2007, S. 12).

„Nicht emotional intendierte Tonhöhenverläufe einer Sprache können als emotional markant in einer anderen Sprache wirken, wodurch Störungen in der Kommunikation hervorgerufen werden können.“

(Nebert 2007, S. 12)

Bei den primären Emotionen aber können grundsätzlich akustische Universalien (Scherer 2000; Hauser 1996) festgestellt und mit dem Konzept der charakteristischen Mimik verglichen werden.

„In den meisten Kulturen sind einige grundlegende primäre Emotionen im Gesichtsausdruck einander so ähnlich, dass sie auf Photographien oder Filmen sofort erkannt werden.“

(Birbaumer/Schmidt 2010, S. 713)

2. Emotionale Lautäußerungen

2.1 Phylogenetische Entwicklung der Stimme

Die Entwicklung der Stimme ist genauso wie die der Emotion an die Evolution der Menschheit gekoppelt. Allgemein lässt sich sagen, dass die vokal-artikulatorischen Ausdrucksmöglichkeiten phylogenetisch entstanden und biologisch determiniert sind (vgl. Stadler Elmer 2005, S. 123).

„Im Verlaufe der Evolution hat der Mensch seine körperlichen Möglichkeiten der Lautgebung zu humanspezifischen Formen wie Sprechen, Singen, Lachen, Weinen usw. differenziert. Diese Ausdrucksformen haben wichtige Funktionen in der Gestaltung des Zusammenlebens der Menschen. Von Geburt an dient die Stimme dazu, mit anderen Menschen in Kontakt zu treten, Wirkungen herzustellen und Befindlichkeiten Ausdruck zu geben. Zuerst geschieht dies durch unkultivierte Signale, dann durch idiosynkratische und intersubjektive Zeichen, und mit fortschreitender Sozialisation werden die in der soziokulturellen Umgebung kollektiv geltenden sprachlichen und musikalischen Lautsymbole und Konventionen integriert.“

(Stadler Elmer 2005, S. 123)

Also kann von der Annahme ausgegangen werden, dass wir primitive emotionale Laute wie sie schon unsere Urahnen ausriefen, verstehen.

Charles Darwin erforscht in seinem Buch von 1872: „The expression of the Emotions in Man and Animals“ (siehe Darwin 1874), wie Affekt auf die vokalen Äußerungen der Tiere wirkt und zieht darin Rückschlüsse auf den Ursprung der menschlichen Expressionen. Er bezeichnet die Stimmorgane bei Mensch und Tier als „wirksame Mittel des Ausdrucks“. (Darwin 1874, S. 84)

Die Entwicklung der Sprache geht einher mit der Genese von anderen Bereichen wie z.B. der musikalischen Entwicklung. Bezüglich der Chronologie differenzieren die Meinungen. Stefanie Stadler-Elmer (2008) meint, dass Sprechen und Singen gleichzeitig entstanden sind, sodass die frühe vokale Kommunikation vermutlich eine Form von Gesang war (vgl. Stadler-Elmer 2008, S. 144). Ebenso befand Darwin diese anfänglichen Lautäußerungen als im Grunde musikalisch.

„[...] dass die Urerzeuger des Menschen wahrscheinlich musikalische Töne ausstießen ehe sie das Vermögen der articulierten Sprache erlangt hatten, und das in Folge hiervon die Stimme, wenn sie in irgend einer heftigen Gemüths-erregung gebraucht wird, durch das Princip der Association einen musikalischen Character anzunehmen strebt.“

(Darwin 1874, S. 88f.)

Andere Studien wie die von Bruhn (1994) besagen wiederum, das Singen hätte sich aus dem Sprechen heraus entwickelt.

„Der Weg zu immer differenzierteren Lautbildungen ist in den ersten Lebensmonaten so eng mit der Entwicklung sprachlicher Fähigkeiten verwoben, so daß angenommen werden kann, daß das Singen phylogenetisch aus dem Sprechen entstanden ist“.

(Bruhn 2010, S. 283)

„Zu der Äusserung vocaler Laute dürften unwillkürliche und zwecklose, in der erwähnten Art und Weise angeregte Zusammenziehungen der Muskeln der Brust und Stimmritze zuerst Veranlassung gegeben haben.“ (Darwin 1874, S. 85) Aus diesen Lautäußerungen heraus sei, so meint Darwin, die Benutzung der Stimme zur Gewohnheit geworden.

„[...] , dass die Stimme, weil sie unter gewissen, Vergnügen, Schmerz, Zorn u. s. w. veranlassenden Bedingungen gewohnheitsgemäss als nützliches Hilfsmittel angewendet worden ist, allgemein gebraucht wird, sobald nur immer dieselben Empfindungen oder Gemüthsbewegungen unter völlig verschiedenen Bedingungen oder in einem geringeren Grade angeregt werden.“

(Darwin 1874, S. 85)

2.2 Frühe Emotionsbedingte Modulation der Stimme

Die Ursache der emotionalen Lautäußerungen ist auch für Darwin (siehe Darwin 1874, S. 86) nicht ganz ersichtlich, doch das sich die Tonhöhe und andere Parameter der Stimme durch Gefühlsregungen verändern, war Darwin schon 1874 gewiss:

„Eine Person, welche sich ruhig über schlechte Behandlung beklagt oder welche unbedeutend leidet, spricht beinahe immer in einem hohen Tone.“

(Darwin 1874, S. 89)

Schon in den Schreien der Babies sind Veränderungen in Grundfrequenz (vgl. Hauser 1996, S. 327) zu vernehmen – vokale Äußerungen, mit denen die Säuglinge Empfindungen signalisieren. Genauso benutzen Erwachsene in ihrer Rhetorik emotionale Modulation, um Aufmerksamkeit zu generieren (siehe Hauser 1996, S. 1) Das der Mensch schon früh zu emotionalen Stimmvariationen fähig ist, beobachtet Darwin bei seinem eigenem, zweijährigen Kind: „[...] das ‚Hm‘ der Zustimmung durch eine leichte Modulation stark emphatisch gemacht wurde, während ein eigenthümlich winselndes Verneinen eine obstinate Bestimmtheit ausdrückte.“(Darwin 1874, S. 87). Diesen Ansatz findet man genauso in der aktuellen Forschung: „Die Integration eines Individuums in die spezifische Art und Weise des Stimmgebrauchs beginnt sehr früh im Leben“ (Stadler Elmer 2005, S. 123). Stadler-Elmer führt weiter aus, dass „eine

Koevolution der kindlichen Ausdrucks-Prädispositionen mit jenen der sozialen Umgebung statt[findet]“. (Stadler Elmer 2005, S. 123)

2.3 Emotionale Laute als körperliche Vorgänge

Beim Ausdruck von Emotionen lässt sich gut die besondere Verbindung von Körper und Stimme beobachten. Vor allem sei auf die wichtige Funktion der Atmung hinzuweisen: „Beim Menschen sind die Respirationsorgane von besonderer Bedeutung beim Ausdruck, nicht bloss bei einer directen, sondern in einem höhern Grade in einer indirecten Art.“ (Darwin 1874, S. 359). Der richtigen Atmung wurde seit jeher eine essentielle Bedeutung zugemessen:

„[...] Vor vielen Jahrhunderten bezeichnete der große römische Redner Seneca den von einer menschlichen Stimme produzierten Ton als *geschlagene Luft*. Ein Larynologe des 19. Jahrhunderts, Gordon Holmes, definierte das genauer: ‚Stimme entsteht aus dem Resultat zweier spezieller Bewegungen: Die eine ursprünglich und lebensnotwendig: die Muskelaktivität; die andere abhängig und materiell: die Bewegung der Luft. Auf diese Beziehung dürfen alle physiologischen Wirkungen der Stimmübungen zurückgeführt werden.“

(Reid 1988, S. 10)

Viele TrainerInnen für Stimmbildung müssen bei ihrer Arbeit mit Menschen verwundert feststellen, dass viele verlernt haben oder gar nicht wissen, wie man richtig atmet. Dabei ist die richtige Atmung die Grundlage des Lebens. Funktioniert sie nicht richtig, dann ist dies sofort in der Stimme zu hören und kann zu einer Reduzierung des Ausdrucks führen. Dies ist alles in Korrelation mit der körperlichen wie auch psychischen Verfassung der/des SprechersIn zu sehen. Einige AutorenInnen weisen daher ganz explizit nur auf die simple Notwendigkeit der Atmung hin (siehe Melaine 2012).

Für die Ausbildung einer Sing- oder/oder Sprechstimme ist neben der richtigen Atmung, auch die Haltung und die Tongebung wichtig. Entspannungs- und Auflockerungsübungen sind unerlässlich bei der Stimmbildung, denn das Verhalten der Muskelspannung wirkt sich auf den stimmlichen Ausdruck aus. Dies ist von jeher ein fester Bestandteil in den meisten Lehr- und Praxisbüchern (siehe Trojan 1955, S. 15ff.). Das, was wir als „Stimme“ wahrnehmen, sind „Schwingungen der Luft [die durch die] Aktivität der Stimmlippen ausgelöst werden“ (Reid 1988, S. 14).

Die Stimme kann weiters durch

„Körperbewegungen [beeinflusst werden], die nicht ursprünglich der Schallerzeugung dienen. Zu diesen heterogenen Merkmalen zählen u. a.: [...] Kaubewegungen des Unterkiefers, der Lippen und Zunge; [...] Orale Zärtlichkeitsgesten; [...] Vorstülpen der Lippen [...]. Zitterbewegungen [...].“

(Trojan 1948, S. 165)

Das Zittern in der Stimme bei Furcht erklärt sich Darwin durch die körperliche Reaktion, das „Erzittern sämtlicher Muskeln“ (Darwin 1874, S. 94). Somit drückt die Stimme auch das aus, was man körperlich fühlt „Als physiologischer Indikator dient die Kontraktion des Lippenheber-Muskels (Musculus levator labii), der deutlich mit dem Ausmaß an Ekelempfinden korreliert.“ (Vaitl 2006, S. 21)

Beim stimmlichen Vorgang sind nicht nur unzählige Muskeltätigkeiten involviert sondern auch die zuständigen Hirnaktivitäten. Spricht man im Zorn so aktiviert sich ein ähnlicher Spannungsaufbau wie bei der Vorbereitung zum Kampf. „In allen diesen Fällen wird Adrenalin aus dem Nebennierenmark freigegeben und Blutzucker aus der Leber ausgeschwemmt“ (Trojan 1955, S. 8). Man kann zusammenfassend sagen, dass alle diese Faktoren einander bedingen:

„Unter ‚stimmlichem Ausdruck‘ wollen wir alle die stimmlichen Mittel verstehen, durch die wir unsere Zustände – als auch den Wechsel von Aufbau- und Leistungshaltung – unseren Mitmenschen kundgeben. Wir sprechen z. B. anders im Zorn als in behaglicher Ruhe.“
(Trojan 1955, S.9)

2.4 Schallbilder emotionaler Laute nach Trojan (1948)

Verbale Äußerungen von Emotionen geschehen also in Begleitung von unterschiedlichen körperlichen wie psychischen Aktivitäten. Trojan (1948) bezeichnet diese Schallbilder als Realisationen von „Akuemen“:

„Im folgenden sei unter ‚Akuem‘ der Inbegriff aller Merkmale verstanden, durch die sich ein Affekt oder Gefühlszustand phonisch und artikulatorisch kundgibt und die sich in den Realisationen wiederfinden müssen, damit deren Bedeutung verstanden werden kann.“
(Trojan 1948, S. 13)

Diese ‚Akueme‘ sind gemäß Trojan (1948) gewisse „Normen [...], nach denen sich jeder Affektausdruck richten muß, um als ein bestimmter, z. B. als der des Zornes oder der Angst, erkannt zu werden.“ (Trojan 1948, S. 15) Schallbilder selbst sind „[...] genetisch älter als die Verständigungsmittel der Sprache.“ (Trojan 1948, S. 36)

Hinsichtlich des Experiments der vorliegenden Arbeit seien hier die Emotionen Ekel, Zorn, Angst, Freude und Trauer in den Worten Trojans als ‚Akueme‘ erklärt. Die ersten, hier beschriebenen, Akueme teilt Trojan (1948) in die Klasse B ein (Trojan 1948, S. 183) und sagt: „Allen Akuemen der Klasse B liegt [...] die Abspaltung eines Objekts vom Subjekt unter Entwicklung von Unlustgefühlen zugrunde.“ (Trojan 1948, S. 184). Zu dieser Gruppe zählt Trojan auch *Ekel*, *Angst* und *Zorn* hinzu. Die vokale Äußerung von Ekel vergleicht er mit dem körperlichen Vorgang des sich Erbrechens und dies sei im Schallbild selbst zu erkennen: „Die Taktgestalt [...] zeigt gleichwohl,

daß das Übel durch impulsive Stöße abgewehrt wird. (Trojan 1948, S. 186). Bei *Zorn* hat Trojan Schwierigkeiten (vgl. Trojan 1948, S. 188ff.) die spezifischen Merkmale zu nennen, die ihn von den anderen Akuemen dieser Gruppe (insbesondere von *Verachtung*) differenzieren, doch sei *Zorn* vor allem an der gesteigerten Dynamik zu erkennen (vgl. Trojan 1948, S. 188). Des Weiteren ist eine Tendenz zur Bruststimme zu beobachten:

„Die Bruststimme zielt auf die Einschüchterung des szenischen Partners. Sie steht im Bunde mit der nach hinten zu verschobenen Artikulationsbasis; der dumpfe Klang wird durch die wohl als Schutz tendenz zu deutende Neigung zu festem Mundverschluß noch gefördert. Überluft und die Dominanz des Konsonantischen verstärken die Geräuschelemente des Schallbilds. Auch das Zittern des Körpers – ein allgemeines Symptom starker Erregung – kann auf die Stimme einwirken und sie eigentümlich verzerren. Alle diese Elemente schließen sich zu einer Gesamtwirkung von gefährlich drohendem Charakter zusammen. Ihr Ziel liegt offenbar darin, im Partner den Affekt der Angst zu erregen, ihn zumindest wehrlos zu machen.“

(Trojan 1948, S. 188f.)

Bei Angst unterscheidet Trojan (1948) zwei unterschiedliche Schallbilder, bei denen es sich wohl „[...] offenbar biologisch verschiedene Reaktionen [handelt].“ (Trojan 1948, S. 190). Das eine Akuem bezeichnet jenes des *Schreckens*, das durch ein entspanntes Schallbild wie beim unbeweglichen Tun äußert (vgl. Trojan 1948, S. 190f.), als würde man versuchen „[...] der Aufmerksamkeit des Gegners durch ein Sich-Ducken, [...] – [zu] entgehen.“ (Trojan 1948, S. 191). Das andere Akuem korreliert genau mit dem Gegenteil, nämlich der „Fluchtreaktion“ (vgl. Trojan 1948, S. 191). Diese nennt Trojan (1948, S. 190) *Furcht* und vermutet dabei Sympathikuserregungen wie bei *Zorn* und *Schmerz* (vgl. Trojan 1948, S. 191). Man erkennt es am „[...] am pulsierenden Atem, am Staccato, am verhärteten Einsatz, an der Dominanz des konsonantischen und des expiratorisch-dynamischen Elements.“ (Trojan 1948, S. 190).

Die nächste Gruppe teilt (u. a. *Freude*) Trojan (vgl. 1948, S. 192) der Klasse C zu und diese Akueme

„[...] bezeichnen zunächst ein Streben nach einem positiven, mit reiner Lust begehrten Wert. Ist dieses Ziel erreicht, so schwingen sie zum Punkt A [...] zurück, um auszudrücken, daß der errungene Wert assimiliert, genossen oder zum Gegenstand einer Einfühlung wird.“

(Trojan 1948, S. 192f.)

Freude habe in der Dynamik Ähnlichkeiten mit dem Schallbild von *Zorn*, wobei die Entladung der Körperkräfte eine andere Ursache hat. (vgl. Trojan 1948, S. 195)

Ein gemeinsames Merkmal der Akueme von Klasse D (Trojan 1948, S. 200) ist die „Entspannung“ (Trojan, S. 200). Das Schallbild der *Trauer* äußert sich vokal nach Trojan (1948) als „Monotonie“ (Trojan 1948, S. 201).

Diese Beschreibungen werden dann in weiterer Folge mit den Ergebnissen der Qualitativen Erhebung (Kap. 5.6, S. 45) verglichen.

3. Gegenwärtiger Stand der Forschung emotionaler Korrelate der Stimme

Dazu gibt es eine Vielzahl von Studien, vor allem im Forschungsbereich der Spracherkennung. Die große Problematik bei der Entwicklung von Emotionserkennungssystemen, ist die Identifizierung der passenden Signalmerkmale zur Erkennung von Emotionen.

Die Forschung im Bereich der Spracherkennung leistet einen wichtigen Beitrag zur Verbesserung von Mensch-Maschine-Dialogsystemen. Vor allem möchte man sich technisch an die natürliche menschliche Kommunikation annähern und sich dieses Wissen bei Computern und anderen Maschinen zu Nutze machen, um dadurch die Effizienz bei gewissen Operationen steigern zu können. Schon jetzt hat ein Tool der Spracherkennung weitläufig den Einzug in unseren Alltag gefunden. Der Spracherkennungsdienst *Siri*, welcher seit dem Modell 4S in iPhones integriert ist, kann Gesprochenes in Text übertragen. Da das Eintippen von Sätzen viel länger als das Einsprechen dauert, bedeutet dieser Fortschritt Zeitersparnis bei der Nachrichtenübermittlung. Ein weiteres Anwendungsgebiet ist die Steuerung von Maschinen mittels Sprache, wie dies zum Beispiel schon in der Automobilindustrie der Fall ist. Doch dabei stößt man unweigerlich auf die Schwierigkeit, dass viele Faktoren, die für das Funktionieren der Kommunikation von Menschen untereinander wichtig sind, beim Dialog zwischen Mensch und Maschine gar nicht oder nicht ausgereift genug vorhanden sind und es somit zu Missverständnissen, Undeutlichkeiten und sonstigen Problemen kommen kann. So kann ein Computersystem bei unterschiedlicher Aussprache oder etwa einer unüblichen Sprechgeschwindigkeit Probleme bei der Identifizierung von Wörtern haben. Bei der zwischenmenschlichen Kommunikation führt auch zusätzliche Information in Form von Mimik, Gestik und Tonfall zu einem besseren Verständnis. Erfolge können die Wissenschaftler genauso bei der Erkennung von Krankheiten durch Stimmanalyse erzielen. So gibt es nach Dr. Max Little und seinem Team eine Technologie, die mit 99% Akkuratessse Parkinson am Telefon erkennen kann (siehe Parkinsonsvoice 1012). Dabei werden u.a. dieselben Parameter wie bei der

Emotionserkennung gemessen (z. B. Tonhöhe, Spektralanalyse, etc.).

Seit etwa 2 Jahrzehnten erkennt man die Notwendigkeit Emotionsforschung bei der Entwicklung neuer technischer Systeme nicht nur zu berücksichtigen, sondern auch daraus Vorteile für die Entwicklung neuer Applikationen und dergleichen zu erzielen. In einem Essay stellt Picard (1995) vom MIT Lab die Idee des „Affective Computing“ als zukunftsweisend dar (siehe Picard 1995). Eingangs postuliert er, dass Computer bald die Fähigkeit haben werden: „Gefühle zu besitzen“ (Picard 1995, S. 1). Danach skizziert er Möglichkeiten der affektiven EDV und erwähnt dabei auch, dass Gefühle in der Wissenschaft bis dato distanziert behandelt wurden, da Emotionsforschung als nicht-wissenschaftlich angesehen wurde. Picard (1995) fasst in diesem kurzen Bericht die essentiellen Komponenten dieses neuen Bereichs zusammen, der „[...] primär die Erkennung und Synthese von Gesichtserkennung, und die Synthese des Tonfalls der Stimme [...]“ (Picard 1995, S. 14) umfasst. Er beschreibt weiters Bereiche, in denen Affektive EDV seinen Einsatzbereich finden könnte: „[...] learning, information retrieval, communications, entertainment, design, health, and human interaction“ (Picard 1995, S. 15). Eine seiner Prophezeiungen, die der „wearable computer“ ist schon längst Realität des modernen Alltags in Form von Smartphones oder Laptops geworden.

Ältere Forschungen zu akustischen Begleiterscheinungen von Emotionen gibt es z. B. von Klaus R. Scherer aus den 1970er Jahren. Scherer forscht sehr viel zu diesem Thema und noch einige weitere seiner Studien werden in dieser Diplomarbeit behandelt. In jener Studie von 1974 (siehe Scherer 1974) hat er mit Hilfe eines Synthesizers emotionale Stimuli modelliert. Er kam zum Schluss, dass es charakteristische Muster akustischer Merkmale bei emotionalen Expressionen (siehe Scherer 1974, S. 108) gäbe. Wie andere vor ihm, erkannte er Beziehungen zwischen Amplitude und Ausdrucksstärke, zwischen Variationen von Tonhöhe, Amplitude, Tempo und der Ausdruckskraft, sowie dass Tonhöhenlevel und Variationen im Verlauf Einfluss auf die Emotionswahrnehmung haben.

Die automatische Erkennung von Emotion in der Mensch-Maschine-Kommunikation könnte in mehrerlei Hinsicht gewisse Arbeitsabläufe erleichtern und findet somit große Aufmerksamkeit in der technischen Wissenschaft. In diversen Studien (Lee/Narayanan 2005; Lee/Narayanan 2003, u. a.) wurde anhand von Audiodateien eines Call-Centers die Erkennung von negativen Emotionen in Dialogen untersucht. Dabei konzentriert sich die Studie auf eine reduzierte Form

einer Trennungsmöglichkeit von Emotionen: negativ vs. nicht-negativ. Mit dieser simplen Unterscheidung könnte z. B. ein automatisiertes Dialog-System optimiert werden (siehe Lee/Narayanan 2005, S. 293): Ist in der Stimme des Anrufers Ärger oder Missfallen zu erkennen, könnte die Maschine den Anrufer automatisch an einen menschlichen Operator weitervermitteln. Bei den akustischen Messungen dieser Studie wurde mit folgenden Parametern gearbeitet (siehe Lee/Narayanan 2005, S. 296): Grundfrequenz (F0), Energie, Dauer und Formanten (F1, F2). Ebenso suchte man in der Studie von Ang *et al.* (2002) nach negativen Emotionen (hier: Ärger und Frustration) im Mensch-Computer-Dialog. Als ein akustisches Resultat wurden die längere Dauer und die langsamere Sprechgeschwindigkeit bei der Emotion *Frustration* angegeben. Ansonsten war für die Emotionserkennung nur die zeitweilig erhöhte Stimme aussagekräftig (siehe Ang *et al.* 2002, S. 4).

Eine Studie, die mit *HMMs* (Hidden Markow Modell, siehe Glossar, Kap. 7 S. 56) *Klassifikatoren* arbeitet, ist eine Forschung zur Erkennung von Emotion auf Basis von Phonem-Klassen (siehe Lee *et al.* 2004). Hier werden Phonem-Klassen etwa *Vokal* oder *nasaler Konsonant* in Zusammenhang mit emotionalen Zuständen gebracht. Ein Ergebnis dieser Studie ist, dass automatische Emotionserkennung Merkmale von Phonemen einbauen muss, da der Typus des Sprachlauts sich auch auf die Formung des Vokaltrakts auswirkt, genauso wie es der emotionale Zustand tut (Lee *et al.* 2004, S. 4). Die Emotionen *Freude* und *Ärger* verwirrten bei der Klassifikation am meisten (siehe Lee *et al.* 2004, S. 4) Die Conclusio fasst weiters zusammen, dass spektrale Merkmale sowie „prosodische Faktoren wie Tonhöhe, Amplitude/Dynamik und Dauer“ (Lee *et al.* 2004, S. 4) signifikante Rollen in der Emotionserkennung spielen (siehe Lee *et al.* 2004, S. 4).

Weiters gibt es Studien, die Bewegungen der wichtigsten Artikulatoren der Stimme: Zungenspitze, unteres Kiefer und die Unterlippe, untersuchen. In einer jüngeren Studie hat Yildirim *et al.* (2005) emotionale Sprachartikulation mit Fokus auf die Vokalproduktion untersucht (siehe Yildirim *et al.* 2005, S. 497). Ein männlicher Sprecher ohne stimmliche Ausbildung sprach 14 Sätze ein, bei denen er die Emotionen *Neutral*, *Ärger*, *Trauer* und *Freude* simulierte. Die meiste Tonhöhenmodulation oder –variation war bei der Emotion *Freude* zu erkennen (siehe Yildirim *et al.* 2005, S. 498). Der Bereich der Unterkieferöffnung ist bei der Emotion *Ärger* aufgetreten (siehe Yildirim *et al.* 2005, S. 500). Der Vokal /IY/ sprach am wenigsten auf emotionale Veränderungen an und darin zeige sich, so Yildirim *et al.*

2005 (siehe S. 500), dass die Artikulation eines Vokals auch den Effekt den eine Emotion auf einen Vokal hat determiniert (siehe Yildirim *et al.* 2005, S. 500).

Bei den Studien, in denen das Alter als Faktor mit untersucht wurde differenzieren die Meinungen. Scherer (1974) behauptet, dass die Altersdifferenzen nicht signifikant waren, dafür die Unterscheidung der Emotionen untereinander viel problematischer seien.

Man benötigt diese Erkenntnisse in der Entwicklung technischer Hilfsmittel, die in rasanter Weise auf den Markt drängen heutzutage mehr denn je.

4. Ausgangspunkt für Pilotstudie

Das folgende explorative Experiment beruht auf dem Prinzip der Basisemotionen. (siehe Kap. 1.2, S. 11). Die Zusammenfassung der Studien und der Literaturrecherche überein (Kap. 2) führt dazu folgende Basisemotionen im Vergleich zum neutralen Zustand zu untersuchen (siehe Kap. 5): Wut, Angst, Freude, Trauer und Ekel.

Die Zusammenfassung der Studien (siehe Kap. 3, S. 19–21) von Scherer 1974, Lee/Narayanan 2005, Ang. *et al.* 2002, Lee *et al.* 2004, Yildirim *et al.* 2005 und jener Studien von Banse/Scherer 1996, Yildirim *et al.* 2004 die im explorativen Experiment behandelt werden ergibt folgende Zusammenstellung von akustischer Parametern zur Analyse der suprasegmentalen (siehe Glossar, Kap. 7 S. 56) Eigenschaften von emotionaler Sprache:

- Tonhöhe (vor allem die Unterscheidung in Hoch u. Tief)
- Tonhöhenbereich (weit oder eng)
- Variationen im Tonfall
- Anzahl der Pausen und die Sprechgeschwindigkeit (schnell oder langsam)
- Verlauf der Grundfrequenz F0 (steigend, gleichbleibende, fallend)
- Stimmhafte/stimmlose Anteile
- Schwankungen in Sprach-Rhythmus, Amplitude und Tonhöhe
- Amplitudenverlauf
- Spektrale Energieverteilung (Artikulation: gepresst, locker...)

Diese Zusammenstellung führt zur Hypothese, dass schon folgende einfach akustische Parameter, wie Grundfrequenz und deren Verlauf sowie Kontur (vgl. Paeschke 2003, S. 32–36) emotionale Aussagen charakterisieren können.

Als Vergleich folgt eine segmentale Analyse der Laute mit dem Hauptaugenmerk auf den Formantbereichen (siehe Anhang 2, S. 70–75). Eine Variable, um **Klangfarbe** der Stimme beschreiben zu können, ist die Lage der Formanten, in insbesondere die „Sängerformanten“ (siehe Sundberg 1987). Es werden F1 und F2 bei den Vokalen analysiert, da man einen Vokal größtenteils schon an den ersten beiden Formanten (F1, F2) erkennen (siehe Lee/Narayanan 2005) kann:

Die beiden ersten Formanten F1 (Eigenfrequenz in der Mundhöhle) und F2 (Eigenfrequenz der Rachenhöhlen) sind charakteristisch für die Vokalfarbe. Physische Kontrolle über die Vokaleigenschaften wird also entweder durch eine gedankliche Vorstellung oder die Formung der Rachenhöhlen und des Mundraums ermöglicht.

(Reid 1988, S. 20)

Es werden die eigenen Ergebnisse mit jenen von Meyer 2004 und Peterson/Barney (1952) (siehe Anhang 2: Teil 2 Vokale) verglichen. In der Annahme das, es die Ergebnisse der eigenen Messungen sich in ähnlichen Formantbereichen wie Meyer 2004 und Peterson/Barney 1952 bewegen.

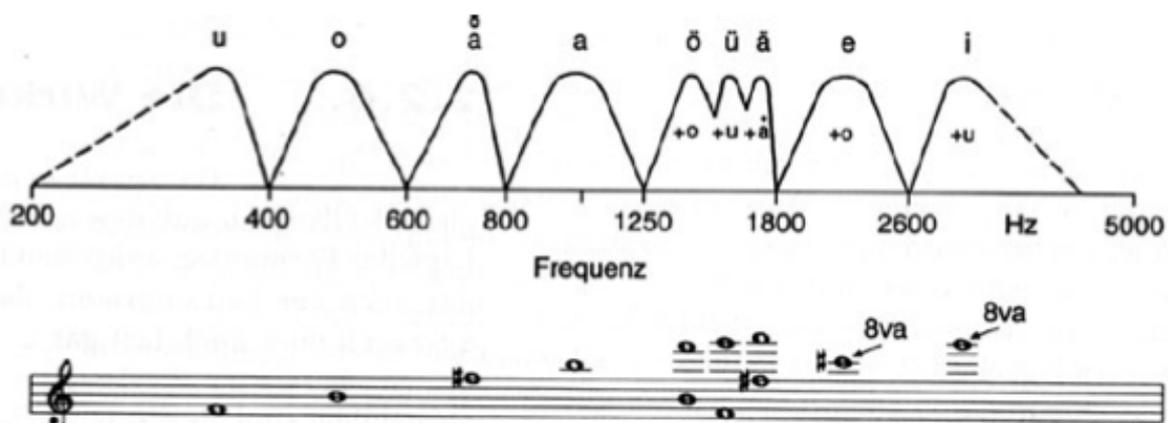


Abb. Vokalformanten (aus: Meyer 2004, S. 33)

Abweichungen davon könnten auf emotionale Modulationen (siehe Lee/Narayanan 2005) hinweisen. Betrachtet man die Studien Lee *et al.* (2004) und Yildirim *et al.* (2005) stößt man auf folgende Fragen:

Welche Vokale einerseits aufgrund ihrer Artikulation prädestiniert für gewisse Emotionen sind? Wie sich andererseits der emotionale Zustand auf die Formung des Vokaltrakts und somit auf die Vokalproduktion auswirkt?

II Experiment

5. Der intendierte emotionale Ausdruck der Stimme und seine akustisch messbaren Merkmale

5.1 Abstract

Die Übersicht (siehe Kap. 4, S. 22) der akustischen Parameter aus den im Kapitel 5 (S. 19) vorgestellten Studien zeigt wie man Emotionen aufgrund ihrer akustischen Charakteristika aus der Stimme entschlüsseln kann.

Das Ziel dieser vorliegenden explorativen Studie ist der Vergleich des intendierten Ausdrucks mit den tatsächlich messbaren akustischen Merkmalen. Bei diesem Experiment wurden die Analyse-Programme „Praat“, „Emapson“ verwendet. Bei den Ergebnissen wurde ein Vergleich mit den Studien Banse/Scherer (1996) und Yildirim *et al.* 2004 (siehe Kap. 5.9, S. 50) vollzogen. Der akustischen Analyse ging eine Rezipienten-Studie voran.

Unter Berücksichtigung der Studien (siehe Zusammenfassung in Kap. 4) kann man sich vorstellen, dass man in diesem vorliegenden Experiment ähnliche Ergebnisse bei F0, F0 Kontur, F0 Range, Energieverteilung und Sprechgeschwindigkeit wie bei Banse/Scherer 1996 und ähnliche – durch Emotionen beeinflusste – Veränderungen bei F1, F2 wie bei Yildirim *et al.* 2004 erhält. In Erwartung multifaktorieller Ergebnisse zeigten die Resultate, dass es viele Übereinstimmungen bei den akustischen Merkmalen von Emotionen gibt. Die Ergebnisse der vorangegangenen Studien konnten bestätigt und um neue Erkenntnisse erweitert werden.

5.2 Einleitung und Theorie

Der Fokus beim folgenden Experiment und der anschließenden akustischen Analyse liegt auf dem tatsächlichen stimmlichen Ausdruck ergo den natürlichen Schallbildern von Emotionen und nicht auf deren synthetische Imitation. Doch Erkenntnisse von Studien der Sprachsynthese lieferten zu dieser Arbeit Grundlagenwissen, da die Definition der akustischen Merkmale des menschlichen Stimmausdrucks eine Vorbedingung für die Sprachsynthese ist.

Ausgehend von der Fragestellung, ob der intendierte stimmliche Ausdruck auch von Rezipienten als ebensolcher aufgenommen wird, wird zu Beginn ein Hörversuch unternommen, um die wesentlichsten Perzepte zu ermitteln (siehe Huber 2005).

Nach der statistischen Auswertung (siehe Sedlmeier/Renkewitz 2008) wird in einer akustischen Analyse untersucht, wie der subjektive emotionale Eindruck auch in messbaren Daten zu deuten ist. Verschiedene Kontrollinstanzen (z.B. Evaluierung und Testdurchläufe vor dem eigentlichen Hörversuch) werden im Ablauf des Experiments eingebaut um mögliche Fehlerquellen gering zu halten.

Hauptsächlich konzentriert man sich in den Studien auf die archetypischen Emotionen, um detaillierte Information über mögliche Abgrenzungen zu gewinnen. (siehe auch Kap. 1.2 und Kap. 4). Bei Yildirim *et al.* (2004), Lee *et al.* (2005 & 2004) wurde nach den akustischen Informationen der Emotionen *Ärger*, *Trauer*, *Freude* und *Neutral* gesucht. In einer Zusammenfassung von mehreren Studien bei Scherer (2000) sind es die Emotionen *Angst*, *Ekel*, *Freude*, *Trauer* und *Ärger*. So werden auch hier die akustischen Eigenschaften der sechs verschiedenen emotionalen Zuständen *Neutral*, *Wut*, *Angst*, *Freude*, *Trauer*, *Ekel* (siehe Kap. 4, S. 22) untersucht.

Die Beispiele werden von zwei Frauen und zwei Männern eingesprochen, die Erfahrung mit dem stimmlichen Ausdruck von Emotion haben. Um die wesentlichsten Aufnahmen zu ermitteln, wird zuerst eine Vorevaluierung mit 3 ExpertenInnen unternommen, danach ein Hörversuch an 45 TeilnehmerInnen unterschiedlicher musikalischer Bildung und demographischen Merkmalen durchgeführt. Um eine möglichst große demographische Reichweite zu erzielen werden Parameter wie idiosynkratische Differenzen, Alter, Geschlecht, u. a. wurden in der vorliegenden Arbeit nicht berücksichtigt.

Bei der Abhörsituation wird darauf geachtet eine Umgebung zu schaffen, wie es sie auch in der Realität gibt. Ganz so als würde man einen Raum betreten, in dem man einer Unterhaltung in einer unbekanntem Sprache zuhört und doch glaubt durch den Klang der Stimmen etwas, zumindest die emotionale Gefühlslage der Unterhaltung, zu verstehen. Um diese Authentizität so weit möglich zu wahren wird eine Art Feldforschung mit einer simplen Gestaltung des Hörtest durchgeführt, da eine Laborsituation sowohl den Ausdruck sowie die Rezeption von Emotionen verfälschen würde.

Das Sprachmaterial soll sich gänzlich vom semantischen Inhalt lösen und so

wurde ein langes Phantasie-Wort kreiert. Ganz nach der Meinung – wie eingangs erwähnt, dass man auch von fremden Sprachen Emotion decodieren kann (Scherer 2000) wird in dieser Studie nur eine Abfolge von Phonemen oder einzelne Vokale verwendet.

Es handelt sich einerseits um ein Replikationsexperiment, da gewisse Teile mancher Studien wiederholt und mit den eigenen Ergebnissen verglichen werden. Andererseits handelt es sich um ein Prüfexperiment, da mehrere Resultate anderer Experimente durch die Erkenntnisse aus der eigenen Studie überprüft werden.

Die Emotionserkennung aufgrund von Sprachsignalen konzentriert sich auf suprasegmentale prosodische Faktoren, also mehr auf die Ausdrucksform als auf den semantischen Inhalt (siehe Banse/Scherer 1996). Die meisten Studien zur Detektion von Emotion in der Sprache analysieren Parameter, die mit Prosodie, Artikulation von Vokalen und der spektralen Energieverteilung verbunden sind. Dabei wird die Tonhöhe, die Dauer von Phonemen oder Silben, die Länge der stimmhaften sowie stimmlosen Regionen, die Dauer von Pausen, die Energieverteilung im Hüllkurvenverlauf, die spektrale Verteilung sowie die ersten 2-3 Formantfrequenzen (siehe Studien von Banse/Scherer 1996; Scherer 1995; Scherer 1974;) analysiert und verglichen.

Generell kann man sagen, dass Differenzen aufgrund von Sprache und Kultur zu einer Verwirrung bei der Klassifikation von emotional gefärbten Expressionen führen können (siehe Nebert 2007). Doch gibt es einen Konsens unter Forschern, dass die Basisemotionen (siehe Kap. 1.2), zu denen auch *Ärger*, *Freude*, *Trauer*, *Ekel*, *Angst* und *Überraschung* zählen, Ähnlichkeiten in allen Kulturen und Sprachen haben (siehe Scherer 2000). Scherer hat in dieser Studie (Scherer 2000) die Möglichkeit der Existenz von einem universalen psychobiologischen Mechanismus von Emotion in der Sprache untersucht, denn fünf Emotionen konnten in neun Sprachen mit einer Genauigkeit von 66% erkannt werden. Hatte Scherer noch in seinen früheren Studien mit elektronischer Inhaltsverdeckung (Rogers *et al.* 1971) durch Stichprobenverbindungen (Scherer 1971) gearbeitet, um zu zeigen, dass eine minimale Anzahl von vokalen Merkmalen ausreicht um emotionalen Ausdruck zu ermitteln, so wurde in einer aktuelleren Studie (Scherer 2000), um einer transkulturellen Forschung dienlich zu sein, ein künstlicher Satz aus verschiedenen Phonemen diverser indo-europäischer Sprachen erstellt, bei dem die Verdeckung des semantischen Inhalts nicht notwendig ist, da keiner existiert. In Anlehnung an

jene Methode wurde bei diesem vorliegenden Experiment als Sprachmaterial auch ein Phantasiewort (siehe S. 30) und Vokale gewählt um den Inhalt zu verdecken. Außerdem werden Assoziationen mit Wörtern, die schon emotionale Konnotationen haben vermieden. Listen solcher emotionaler key-words, die als korrelative Information zwischen der Emotion und einer spezifischen Aussage fungieren, findet man u.a. in englischer Sprache bei Plutchik (1994). Bei den anderen erwähnten Studien, besteht das Sprachmaterial zumeist aus Sätzen mit semantischem Inhalt.

Einige Studien arbeiten nach dem Konzept des 3-dimensionalen emotionalen Raums, in dem die Kategorisierung von Emotion nach den Grundformen Valenz, Aktivierung und Dominanz passiert (siehe Grimm *et al.* 2007; Ilie/Thompson 2006). Hauptgegenstand dieser Diplomarbeit sind jedoch Studien, die den kategorialen Zugang gewählt haben. Daher wurde der kategorielle und nicht der dimensionale (Valenz – Erregung) Zugang gewählt. Somit werden sich im weiteren Verlauf der Arbeit keine Daten zu Valenz oder Erregung finden.

Folgende Ergebnisse werden im Detail mit den Ergebnissen des Experiments dieser Diplomarbeit verglichen:

In der Studie von Banse/Scherer (1996) wird eingangs die Zusammenfassung der Ergebnisse mehrerer Studien, die wiederholt Forschung zu den emotionalen Schallbildern von Ärger, Angst, Trauer, Freude und Ekel betrieben haben, gezeigt:

Anger: [...] generally seems to be characterized by an increase in mean F0 and mean energy. [...] increases in high-frequency energy and down-ward-directed F0 contours. The rate of articulation usually increases.

Fear: There is considerable agreement on the acoustic cues associated with fear. [...] showing increases in mean F0, in F0 range, and high-frequency energy. Rate of articulation is reported to be speeded up. An increase in mean F0 has also been found for milder forms of the emotion such as worry or anxiety.

Sadness: As with fear, the findings converge across the studies that have included this emotion. A decrease in mean F0, F0 range, and mean energy is usually found, as are downward-directed F0 contours. There is evidence that high-frequency energy and rate of articulation decrease. [...]

Joy: [...], we find a strong convergence of findings on increases in mean F0, F0 range, F0 variability, and mean energy. There is some evidence for an increase in high-frequency energy and rate of articulation.

Disgust: [...], the results for disgust tend to be inconsistent across studies. [...]. The studies that have used the former found an increase in mean F0, whereas those that have used the latter found the reverse – a lowering of mean F0. This inconsistency is echoed in the decoding literature.“

(Banse/Scherer 1996, S. 616)

Anschließend führten sie eine weitere aufwendige Studie zu 14 verschiedenen Emotionen durch, wo sie z. B. eine Unterscheidung zwischen „hot anger“, „panic fear“, „cold anger“ usw. (siehe Banse/Scherer 1996, S. 9) getroffen haben. Diese

Ergebnisse sind mit dem vorliegenden Experiment nicht vergleichbar, da es sich nur mit den oben schon genannten sechs primären emotionalen Zuständen befasst.

Bei Yildirim *et al.* (2004) wurde untersucht wie die Sprache moduliert, wenn sich die Emotion vom neutralen Status zu einem gewissen emotionalen Zustand ändert. Dazu wurden die stimmlichen Äußerungen zu den Emotionen Trauer, Ärger, Freude und Neutral, gesprochen von einer Schauspielerin, aufgenommen. Es wurden Messungen der Prosodie, Vokalartikulation und spektralen Energieverteilung durchgeführt. Bei der Dauer zeigte sich eindeutig, dass Trauer, Ärger und Freude längere Durchschnittswerte haben als Neutral (siehe Yildirim *et al.* 2004, S. 2). Ebenso erscheinen die Ergebnisse der **Vokale** signifikant:

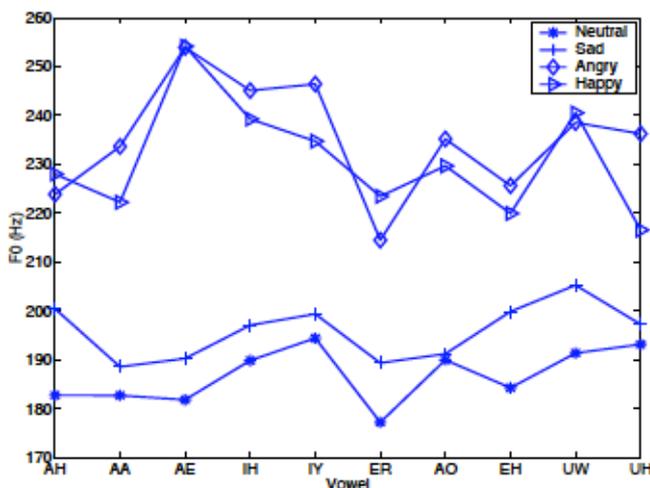


Abb.: FO Vokale (aus: Yildirim *et al.* 2004, S. 3)

Diese Abbildung zeigt die Grundfrequenz (F0) der **Vokale** in der jeweiligen emotionalen Kategorie. Es wurde beobachtet, dass die F0 der Vokale bei der neutralen Sprache tiefer ist als bei den anderen emotionalen Stadien. Weiters stellte sich heraus, dass Ärger/Freude und Trauer/Neutral ähnliche F0 Durchschnitte haben.

Yildirim *et al.* (2004) erkennen einen wesentlichen Einfluss von Emotionen die **Formatfrequenzen (F1 und F2)** bei den Vokalen und erklären dies mit der unterschiedlichen Zungenstellung.

"Interestingly, difference in the formant patterns between the two groups of emotion (i.e., anger/happiness and sadness/neutral) are better reflected in back vowels such as /a/ than in the front vowels in this speaker. Difference in the manipulation of the lip opening and/or the tongue positioning at the rear part of the vocal tract could be underlying factors. It may be noted that we can't draw any conclusion on the variability of formant frequencies as a function of emotion as they vary depending on which formant is considered. For instance, the sad speech shows the smallest variability for F1, but it is the happy one for F2."
(Yildirim *et al.* 2004, S. 3)

Die folgenden zwei Darstellungen zeigen die durchschnittlichen Frequenzen von F1 (Abb. links) und F2 (Abb. rechts):

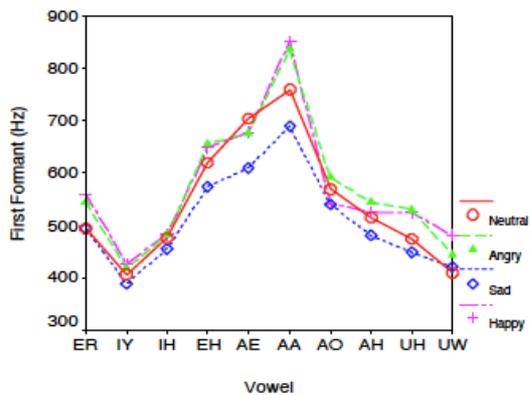


Abb. F1 Vokale (aus: Yildirim *et al.* 2004, S. 3)

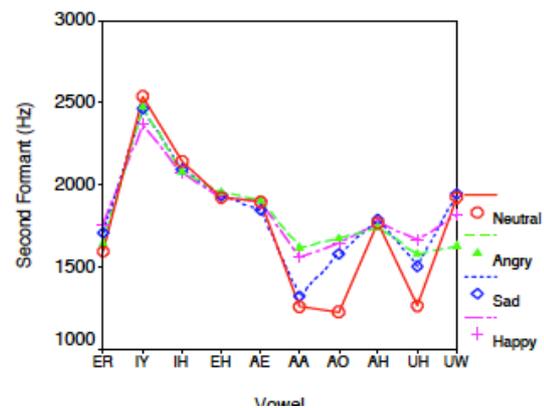


Abb. F2 Vokale (aus: Yildirim *et al.* 2004, S. 3)

Bei der Sprecherin (Yildirim *et al.* 2004) zeigen sich ähnliche akustische Eigenschaften bei den Emotionen *Freude/Ärger* und *Neutral/Trauer*. Verärgerte oder freudige Sprache ist durch eine längere Ausdrucksdauer, kürzeren Pausen, einer höheren Tonlage und einen weiteren Energiebereich charakterisiert, welche auch die Merkmale von übertriebenen oder hyperartikulierte Sprechen sind (siehe Yildirim *et al.* 2004, S.1). Die Emotion Trauer scheint sich durch die Stille zwischen den Wörtern und dem langsamen Tempo von anderen Emotionen zu unterscheiden. Es wurde bei Trauer eine etwas niedrigere Tonhöhe (pitch) mit einem etwas engeren Tonhöhenbereich als bei Neutral festgestellt (siehe Yildirim *et al.* 2004, S. 1). Die Trennbarkeit innerhalb der Gruppen erwies sich als schwierig und Yildirim *et al.* (2004) sind der Ansicht, dass die in ihrer Studie verwendeten konventionellen akustischen Parameter nicht ausreichend genug seien. Doch erkannte man die Differenz innerhalb der Gruppen am besten in den hinteren Vokalen wie [a]. Das könnte mit der größeren Lippenöffnung und/oder einer engeren Zungenstellung am weichen Gaumen zusammen hängen (siehe Yildirim *et al.* 2004, S. 1).

5.3 Methode

Sprachmaterial

Das zu analysierende Material besteht aus den fünf Vokale (/a/, /e/, /i/, /o/, /u/) und dem lautpoetischen Wort *Pajuatükonäschmupfilers* jeweils einmal in den 6

Basisemotionen stimmlich ausgedrückt: *Neutral, Wut, Freude, Angst, Trauer und Ekel*. *Wut & Zorn* ist zumindest literarisch gesehen eine Steigerung von Ärger (siehe Studien Banse/Scherer 1996, Yildirim *et al.* 2004).

Zum lautpoetischen Wort *Pajuatükonäschmupfilers*

Dies wurde nach phonetischen Prinzipien des indogermanischen Sprachstamms kreiert: Es sind alle Vokale (oral und stimmhaft) vorhanden: *a i e o u*

Infolgedessen sind einige Monophthonge (Einzelvokale) einschließlich den Umlauten /ü/ und /ä/ enthalten.

Weiters ist ein Diphthong (siehe Glossar, Kap. 7, S. 56) zu finden: /ua/

Damit die Aussprache nicht zu kompliziert wird, sollte das Wort nicht zu lange sein und deshalb beinhaltet alle Konsonanten sondern nur Vertreter der einzelnen Konsonanten-Gruppen:

- 1.) Verschlusslaute: *Plosive: /p/; *Nasal: /m/ und /n/ *Vibrant: /r/
- 2.) Engelaute: *frikative: /f/ und /s/
- 3.) Approximant: /j/
- 4.) Lateral: /l/
- 5.) Affrikative: /pf/

Wobei Vokalquantität (Länge der einzelnen Vokale) sowie Qualität (Artikulation) den SprecherInnen obliegt. Ob z. B. e nach dem Nasallaut n als sogenanntes Murrellaut, „Schwa“, [ə] artikuliert wird, ist dem/der jeweiligen SprecherIn überlassen. Es befinden sich im Wort 12 stimmhafte und acht stimmlose Laute.

Aufnahme Vokalistin

Alle Stimmaufnahmen wurden mit einem Kondensator-Großmembran Mikrofon 4400a der Firma SE Electronics mit Nieren-Richtcharakteristik und folgenden Einstellungen aufgenommen: dB 0, Low Cut, Phantomspannung 48 Volt und keine Automationen.

Als Recorder fungierte ein Zoom H4n. Die Audio-Files sind im Wave-Format mit der Sample Rate: mind. 44,1 kHz, 24 Bit.

Die Ton-Aufnahmen erfolgten in der alltäglichen Umgebung (Arbeit oder Zuhause) der jeweiligen Sprecherin oder des jeweiligen Sprechers. Es wurde darauf geachtet, dass sich absorbierende Flächen im Raum befanden. Es war eine bewusste Entscheidung der Versuchsleiterin, die Aufnahmen nicht in einem Tonstudio

vorzunehmen, denn eine Studiosituation könnte zu einer gewissen Künstlichkeit (Vokalist*innen ‚bemühen‘ sich mehr, etc.) führen, was bei diesem Experiment nicht erwünscht ist. So kam es aber zu kleinen akustischen Unterschieden (z. B. unterschiedlicher Raumklang).

Die 4 Personen, semi-professionell und professionell, sind im stimmlichen Emotionsausdruck geschult und haben Erfahrung darin.

Der erste Sprecher, im weiteren **„Sprecher 1“** genannt – ist ein semi-professioneller Tenor und 34 Jahre alt. Er besitzt eine langjährige Gesangsausbildung und hat unter anderem am Konservatorium Wien studiert. Seine Muttersprache ist Spanisch, doch spricht er fließend Deutsch. Im Hauptberuf ist er Designer und arbeitet darüber hinaus als Sänger und Tänzer.

„Sprecherin 2“ ist semi-professionelle Stimmkünstlerin (Hauptberuf DJ, Model, Musikerin) und 33 Jahre alt. Sprecherin 2 ist die VI dieses vorliegenden Experiments. Sie verfügt über verschiedene musikalische und stimmliche Ausbildungen. Sie hat u. a. verschiedene Projekte mit internationalen Lautpoeten geleitet. Sie kreiert eigene Performances im Genre der akustischen Kunst mit Konzentration auf Lautpoesie und Ton-Installationen. Vor der Aufnahme absolvierte sie Aufwärmübungen für Stimme und Körper von etwa einer halben Stunde.

„Sprecher 3“ ist vom Hauptberuf Schauspieler, 46 Jahre alt und eignete sich in langjähriger musikalischer und stimmlicher Ausbildung eine hohe Professionalität an. Er ist erfolgreicher Bühnenschauspieler, inszeniert darüber hinaus seine eigenen Stücke und bietet Unterricht in Schauspiel und Sprechen an. Vor der Aufnahme wärmte er ungefähr 2 Minuten seine Stimme und die nötige Muskulatur auf.

„Sprecherin 4“ ist professionelle Sängerin, Komponistin, off-voice und unterrichtet Gesang. Sie ist 36 Jahre alt und gerade mit einem neuen Album auf Tournee. Von Kindheit an hatte sie Musikunterricht (Gitarre, Klavier) und absolvierte an der Universität Graz das „Konzertfach für Jazzgesang“. Da sie an diesem Tag mehrere Gesangsübungen durchgeführt hatte, war ihre Stimme bei der Aufnahme sehr gut aufgewärmt.

Evaluierung Experten

Da jeweils 2-3 Versionen pro Beispiel aufgenommen wurden, musste aus dem gesamten aufgenommenen Material die Perzepte für die nächste Stufe der Evaluierung ausgesucht werden. Die VI wählte gemeinsam mit einer amerikanischen

Sängerin/Gesangslehrerin und einem Schauspielerschüler, der zusätzlich über eine Radiosprecherausbildung verfügt, dieselbe Anzahl an stimmlichen Ausdrücken für alle SprecherInnen und Emotionen. Man einigte sich darauf die zusätzlich aufgenommenen Emotionen *Überraschung* und *Scham* für dieses Experiment wegzulassen. Einerseits weil ein Vergleich mit den oben genannten Studien Banse/Scherer (1996), Scherer (2000) und Yildirim *et al.* (2004) gezogen werden soll. Andererseits weil die Verwechslungsgefahr dadurch verringert wird und es infolgedessen zu aussagekräftigen Ergebnissen kommt. *Scham* ist eine Emotion, die schwer abzugrenzen ist (vgl. Tiedemann 2007, S. 16–23). Eine Differenzierung zu *Angst* ist bei dieser Emotion schwierig (vgl. Tiedemann 2007, S. 23f.; Plutchik 2001, S. 344). Es ist geplant, diese zwei Emotionen in einer Folgestudie als Vergleich heranzuziehen.

Demzufolge ist im Hörversuch das lautpoetische Wort pro SprecherIn in allen 6 emotionalen Zuständen jeweils einmal vorhanden. Die Vokale /a/, /e/, /i/, /o/ und /u/ sind genauso in allen 6 Gefühlslagen von jedem/r der 4 Spr. vorhanden. Das ergibt 144 Beispiele.

Test

Die Problemstellung, dass Laien aller Altersgruppen den affektiven Zustand des/der SprecherIn unabhängig vom semantischen Inhalt der Äußerung beurteilen sollen bildete die Grundlage für die Erstellungsweise des Hörversuchs.

Der Test selbst wurde im Programm *Ableton Live* erstellt. Dabei wurden die Reihenfolge der Beispiele randomisiert arrangiert. Das Arrangement (Reihenfolge der Beispiel) bleibt für alle Vpn gleich. Eine angenehme Lautstärke für den jeweiligen Hörer ist individuell am Anfang des Testes einzustellen.

Eine weibliche Stimme führt durch den Test. Am Anfang sind 6 Hörbeispiele, bei denen entweder ein Vokal oder das lautpoetische Wort jeweils in einem der 6 emotionalen Zuständen zu hören ist. Der jeweilige emotionale Zustand der Beispiele wird aber nicht angegeben. Dieser erste Durchgang dient lediglich dazu, die Rezipienten auf den weiteren Verlauf vorzubereiten, damit sie nicht bei den ersten Hörbeispielen des Versuchs in irgendeiner Weise überrascht sind, denn das könnte das Testergebnis verfälschen. Nach Ablauf der 6 Beispiele beginnt der eigentliche Test. Da die gerenderte Datei einfach in *itunes* abgespielt wird und nach Drücken des „Play“-Knopf durch die VI von Beginn bis zum Schluss ohne Unterbrechung

durchläuft, kann sich der Hörer nur auf die auditive Wahrnehmung konzentrieren.

Weiters ist der Evaluierungsbogen, in dem man die Emotion, die man pro Beispiel empfindet und ankreuzt, sehr einfach gestaltet. Somit ist der Testablauf so generiert, dass auch Menschen, die keine Erfahrung mit Computer haben, diesen auch nicht benutzen müssen. Es werden damit Situationen vermieden, in denen sich die Vp nicht wohl fühlt, da sie/er sich mit dem Umgang von Technik nicht vertraut fühlt. Auch dies könnte sich sonst negativ auf die Ergebnisse auswirken, da es für eine Erhebung zu emotionalen Zuständen essentiell ist, dass die Vpn entspannt sind.

Um den Testablauf optimal zu gestalten wurde vor dem eigentlichen Testbeginn noch eine Evaluierung mit zwei weiblichen Hörerinnen und einem männlichen Hörer in einem kleinen **Vorexperiment** vollzogen.

Danach wurde der zeitliche Ablauf des Test wurde dann randomisiert und gestaltet sich pro Klangbeispiel wie folgt: 1 sec. vor Klangbeispiel, 4 sec. nach den Vokalen, 3,5 sec. nach dem lautpoetischen Wort. Diese Pausen reichten aus, um den Teilnehmer das Ankreuzen der Emotion in aller Ruhe zu ermöglichen. Sie waren aber auch nicht zu lange gewählt, denn es sollte ein gewisses intuitives Handeln bewahrt bleiben, das durch zu lange Zeit des Überlegens verfälscht werden könnte. Mit den einführenden Beispielen und der auditiven Testanleitung ist der Test insgesamt 17 min. 36 sec. lang.

Die Examinierung ist per Kopfhörer (Sennheiser HD 265 linear) damit dieselbe Abhörsituation für alle Testpersonen gewährleistet ist. Alle Vpn hören nur einen Durchlauf. Alle VersuchsteilnehmerInnen bekommen vorm Testbeginn eine persönliche Instruktion durch die Versuchsleiterin, um Ihnen die Thematik (paraverbale Sprache, etc.) und die Sinnhaftigkeit des lautlichen Versuchsmaterials näher zu bringen.

Da das Hauptaugenmerk auf der anschließenden akustischen Analyse liegt, ist die Variablenselektion für die Zuhörerschaft äußerst reduziert. Die Variable der musikalischen und stimmlichen Ausbildung war für die Charakterisierung der Vpn aber nicht unbedingt für das Resultat selbst wichtig.

Für den Großteil der Vpn erfolgte die Testung im Musikraum der Versuchsleiterin, der sich wie ein gemütliches Wohnzimmer gestaltet. Es wurde dabei auf eine heimelige, natürliche Umgebung geachtet (z. B. Tee und Kuchen serviert). Bei 15 Leuten erfolgte die Testung zuhause in ihrer natürlichen Umgebung. Dieselben standardisierten Rahmenbedingungen (Kopfhörer, usw.) galten aber für alle Vpn.

Während der Testung war die/der Vp ganz für sich alleine und ungestört. Also passierte jeder Durchgang als Einzelversuch.

Alle Vpn wirkten freiwillig, also ohne Bezahlung mit.

Die zusätzlichen, schriftlich zu beantwortenden Fragen am Ende des eigentlichen Tests waren:

**Welche Emotion war deiner/ihrer Meinung nach am besten zu erkennen? Woran lag das?*

**Welche Emotion war am schwierigsten zu erkennen? Warum?*

**Welche Emotionen sind am schwierigsten zu unterscheiden? Warum?*

**Welche Emotionen grenzen sich am deutlichsten voneinander ab? Wodurch?*

Die Beantwortung dieser Fragen war nicht obligat. Diesbezügliche Ergebnisse wurden für eine qualitative Erhebung heran gezogen.

5.4 Ergebnisse Evaluierung der Perzepte – Rezipienten-Studie

Die Ergebnisse der einzelnen Hörversuche wurden in ein Excel-Blatt übertragen und mittels des Statistikprogramms SPSS ausgewertet. Es sind rein deskriptive Statistiken angewandt worden.

Geschlecht der Versuchspersonen

		Geschlecht			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	m	18	40,9	40,9	40,9
	w	26	59,1	59,1	100,0
	Total	44	100,0	100,0	

Abb. Geschlecht Vpn

Bei der Auswertung der Test-Ergebnisse musste eine Teilnehmerin von der Liste gestrichen werden. Obwohl sie dieselben Instruktionen wie alle anderen vor Testbeginn erhielt, hat sie mehrere Male zwei statt nur eine Emotion pro Beispiel angegeben. Somit ist ihr Testbogen ungültig und die Anzahl der Rezipienten reduzierte sich dadurch auf 44 Personen. Davon sind 26, also mehr als die Hälfte (59, 1 %) weiblich und 18 Personen männlich.

Alter der Versuchspersonen

Das Durchschnittsalter der TeilnehmerInnen liegt bei 35 Jahren, mit einer Standardabweichung von 13,56%. Die jüngste Teilnehmerin ist 11, der älteste Teilnehmer ist 70 Jahre alt.

		Statistic
Alter	Mean	35,45
	Median	32,00
	Std. Deviation	13,562
	Minimum	11
	Maximum	70
	Range	59
	Interquartile Range	19

Abb. Alter Vpn

Es wurde eine ziemlich große Altersreichweite von 59 Jahren erzielt.

Berufsgruppen

Weiters wurde auch eine breite Palette an Berufsgruppen erreicht. Einige Beispiele für Berufe, wo jeweils eine Person unter den VersuchsteilnehmerInnen vertreten ist: Angestellte, Arzt, Autorin, Gesangslehrerin, Mutter, Wissenschaftlerin, Yoga-Lehrerin u. a. Bei folgenden Berufen waren 2 oder mehrere VertreterInnen: Architekt, bildende Künste, Schülerinnen, selbstständig (aus unterschiedlichen Bereichen wie z. B. Pharma oder Film), Models (männlich und weiblich), Pensionisten, Landwirte, u. a. Die Gruppe von StudentInnen war mit 8 Personen am größten.

Muttersprache der Versuchspersonen

84,1 gaben Deutsch als ihre Muttersprache an. Alle anderen (3 Französisch, 1 Kroatisch, 1 Persisch, 2 Slowakisch) gaben Deutsch fließend und als ihre Zweitsprache an. Alle 44 Personen sprechen Englisch, mehr als die Hälfte noch eine oder weitere Sprachen.

Deutsch	37
Französisch	3
Kroatisch	1
Persisch	1
Slowakisch	2
Total	44

Abb. Muttersprache Vpn

Musikalische Ausbildung der Versuchspersonen

30 Personen verfügen über eine musikalische Ausbildung. Bei 24 sind es 3 Jahre oder mehr. Bei 3 Personen sind es 10, bei einer Person sogar 19 Jahre. Wenn man die TeilnehmerInnen mit einer musikalischen Ausbildung von 1 – 2 Jahren zu jenen mit keinem musikalischen Unterricht hinzurechnet, dann ergibt sich eine gute Balance von Laien (20 Personen) und Experten (24 Personen).

Stimmliche Ausbildung der Versuchspersonen

Mehr als die Hälfte gaben keine stimmliche Ausbildung, welche nicht nur den Unterricht in Gesang sondern auch Chorerziehung oder das Stimmtraining in beispielsweise Rhetorik oder Schauspiel bedeutet, an. 12 Personen gaben ein Training von unter 2 Jahren an. Nur 10 Personen hingegen verfügen über eine stimmliche Erziehung von 3 oder mehreren Jahren bedeutet. 3 Rezipienten, die um einiges älter als der Durchschnitt sind, gaben Ausbildungszeiträume von 17, 20 und 25 Jahren an.

Ergebnisse Sprecher

Der intendierte emotionale Ausdruck wurde bei Sprecher 1 von 63, 7 %, bei Sprecherin 2 von 76, 7 %, bei Sprecher 3 von 57, 6 % und bei Sprecherin 4 von 71,5 % der HörerInnen auch als solcher wahrgenommen. Der *neutrale* Ausdruck wurde bei den zwei professionellen SprechernInnen (3 und 4) mehr als etwa 10 % besser als bei den zwei anderen, den semi-professionellen, SprechernInnen verstanden.

Sprecher 1 konnte am besten *Neutral* mit 76,3 % und *Wut* mit 69,3% richtiger Bewertung übermitteln. Probleme hatte Sprecher 1 beim stimmlichen Ausdruck von *Freude* denn mehr als die Hälfte (51,9%) gaben hier eine andere Emotion an.

Sprecherin 2 brachte am besten die Emotionen *Freude* (98,1% richtige Treffer) und *Ekel* (92,8%). *Neutral*, *Wut* und *Trauer* hielten sich bei dieser Sprecherin die Waage (70,3% – 76,2%). Bei *Angst* konnten mehr als die Hälfte (53,6 %) die evozierte Emotion nicht erkennen.

Bei **Sprecher 3** wurde nach der Emotion *neutral* die *Wut* am besten verstanden (71,5%). *Angst*, *Freude* und *Trauer* wurden bei diesem Sprecher nur von 46 – 56,2%, also von ungefähr der Hälfte, als die intendierte Emotion wahrgenommen. *Ekel* wurde von 63, 1 % nicht als solches empfunden.

Bei der **Sprecherin 4** wurde *Wut* genauso gut verstanden wie *Neutral* (88,5 %), wie auch die Emotion *Freude* von 84 % richtig erkannt wurde. *Ekel*, *Angst* und *Trauer* wurde von ein wenig mehr als der Hälfte (56, 2 – 58, 8 %) als solche begriffen.

Perzepte

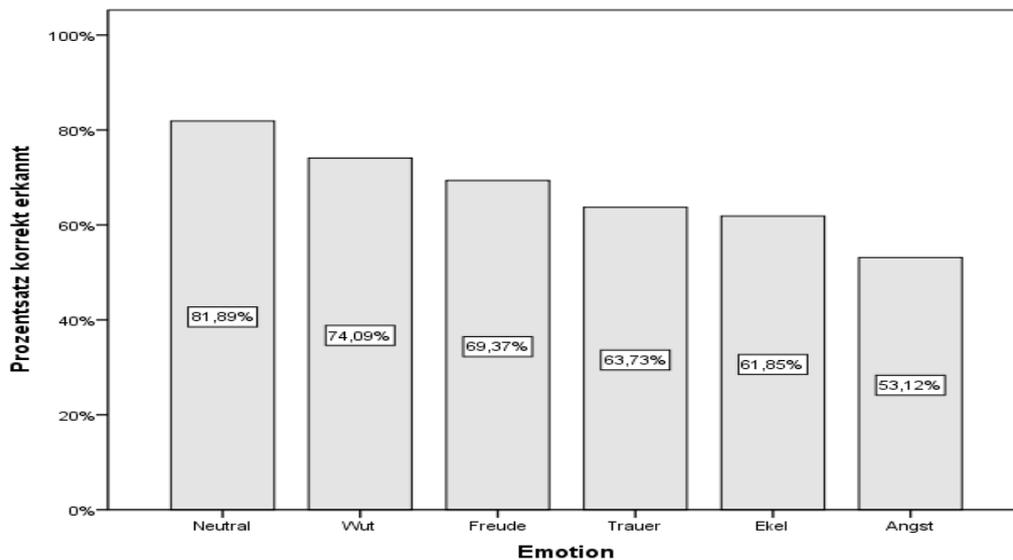


Abb. Prozentsatz der korrekt erkannten intendierten Emotionen

Diese Abbildung oben bietet eine Übersicht über die Erkennungsrate von allen untersuchten Emotionen. *Neutral* wurde mit 81,89%, vor *Wut* (74,09%) und *Freude* (69,37%) am besten erkannt. Das Gesamtergebnis der Erkennung aller Emotionen liegt bei 67,4% Akkuratessse. Bei 6336 möglichen Treffern (144 Beispiele x 44 Versuchspersonen) erfolgte 6277mal eine Bewertung. Somit sind 59 Möglichkeiten ausgelassen worden. Bei den Bewertungen waren 4228 Angaben (67,4%) richtig (siehe Abb. unten).

Total	Count	2049	4228	6277
	% within Korrekt	32,6%	67,4%	100,0%

Abb. Gesamtergebnis Trefferquote

Ingesamt wurden 67,4 % der emotionalen Intention auch als solche erkannt.

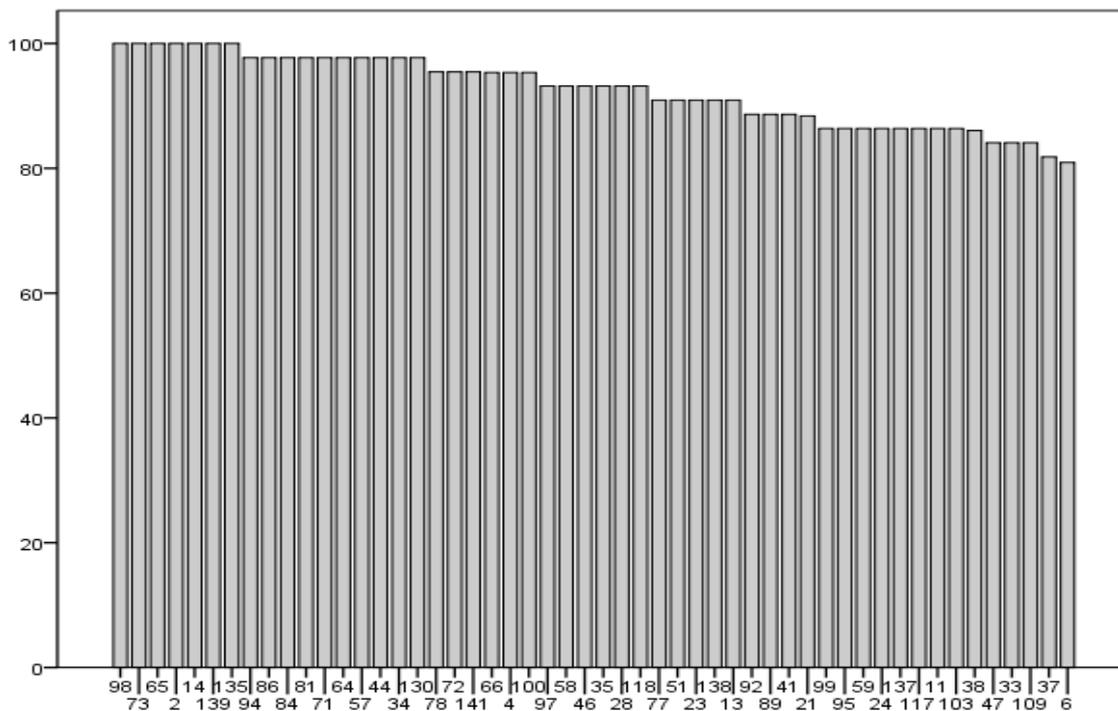


Abb. 52 Beispiele mit der höchsten Trefferquote

Die Abbildung oben zeigt die 52 Beispiele mit der **höchsten Trefferquote**. Unter den Top 20 (95,5 – 100 % Genauigkeit) befinden sich 7 Beispiele mit der stimmlichen Äußerung von *Ekel*, 4 *neutral* gesprochene Vokale, 3 *wütend* geäußerte lautpoetische Worte und 6 Beispiele mit der Intention *Freude*. Es befinden sich keine Beispiele mit den Emotionen *Trauer* oder *Angst* darunter.

Ekel wurde bei Beispielen 73, 137, 35, 139, 130, 78, 81 mit den Vokalen /i/ oder /e/ besser als mit den 3 anderen Vokalen erkannt und sind unter den 52 Beispielen mit der höchsten Erkennungsrate im Gesamtergebnis. *Ekelig* geäußerte /i/ sind von Sprecher 1, 3 und 4 sind unter den ersten 20. Ebenso ist das /i/ dieser Emotion von Sprecherin 2 unter den 30 am besten erkannten Beispiele. Die Erkennungsrate bei diesen Beispielen der Emotion *Ekel* liegt bei 86,4 – 100 %. Nur ein einziges Beispiel (Nr. 70), bei dem /e/ als *Ekel* geäußert wurde, ist von nur 22,7 % auch als die intendierte Emotion wahrgenommen worden. 6 Beispiele mit dem stimmlichen Ausdruck von *Ekel* landeten überhaupt unter den 20 Beispielen, die am besten erkannt wurden. Darunter auch das lautpoetische Wort mit der Emotion *Ekel* von Sprecherin 2 gesprochen (Bsp. 2), welches zu 100 % erkannt worden ist. Bei den Sprechern 1, 3 und 4 war die Rate durchschnittlich. Da *Ekel* bei allen Beispielen mit den Vokalen /u/, /a/ und /o/ unterdurchschnittliche Ergebnisse erzielte, ist das Gesamtergebnis *Ekel* nur bei 61,85%. Hierbei bildeten nur die Beispiele 98 und 118

der Sprecherin 2 die Ausnahme, wo bei ersterem der Vokal /a/ 100 % akkurat und beim zweitem der Vokal /u/ mit 93,2 % richtig erkannt wurde. Auch beim Vokal /o/ wurde bei dieser Sprecherin das Gefühl *Ekel* mit 69,8% deutlich besser verstanden, als bei allen anderen Sprechern, wo es sich die Ergebnisse um einen Mittelwert von 32 % bewegen.

Die **neutralen** Ausdrücke mit der höchsten Trefferquote (95,5 – 100%) waren die Vokale /a/ und /i/, gesprochen von Sprecher 3, der Vokal /e/ jeweils einmal von Sprecherin 4 und einmal vom Sprecher 1 geäußert. Das erste *neutral* geäußerte /u/ liegt bei der Verständlichkeit weiter hinten (88,6%) und stammt von Sprecherin 4. Dieses wurde nur 5 x mit Trauer verwechselt. Beim Beispiel 101, wo /u/ in *neutraler* Weise von Sprecherin 2 geäußert wurde, ist diese Gefühlslage von 11 Personen mit Trauer verwechselt worden. Beim Sprecher 1 herrschte beim neutral formulierten Vokal /u/ Verwirrung, denn dieser wurde mehrere Male mit *Freude*, 1x mit *Wut* und 1x als *Ekel* wahrgenommen. Beim Sprecher 3 wurde das intendierte *neutrale* /u/ mit der Trefferquote von 68,2 % ein paar Mal *ekelig* oder *freudig* und 1x als *wütend* empfunden. Es befindet sich kein einzig *neutral* gesprochenes /o/ unter den 52 Beispielen mit der höchsten Trefferquote. /o/ in *neutraler* Form wurde beim Sprecher jeweils 4x mit *Trauer*, *Angst* oder *Freude* verwechselt. Bei Sprecherin 2 wurde das *neutrale* /o/ einige Male als *Trauer* oder *Angst* und 2 x als *Freude* wahrgenommen. Beim Sprecher 3 lag die Verwechslung bei *Neutral* jeweils 3x bei *Trauer* und *Angst*, bei Sprecherin 4 wurde statt *neutral* mehrere Male *Trauer*, 1x *Angst* und 1x *Freude* angegeben. Das *neutral* von Sprecherin 4 geäußerte Wort wurde von 93,2% auch als solches erkannt.

Die Vokale /u/ und /o/ in **freudiger** Weise von Sprecherin 2 gesprochen, wurden auch zu 100 % als die intendierte Emotion verstanden. Bei 97,7 % der Rezipienten evozierte das lautpoetische Wort (Bsp. 94), die Vokal /a/ und /e/ von Sprecherin 2, und der Vokal /o/ von Sprecherin 4 als *Freude* geäußert auch diese Emotion. Im übrigen wurde das *freudige* lautpoetische Wort von Sprecher 3 (Bsp. 4) und der Vokal /i/ von Sprecherin 2 von 95,3% auch als *Freude* vermerkt.

3 **wütend** übermittelte lautpoetische Wörter (Bsp. 71, 64, 44) von Sprecher 2, 3, und 4 wurden von 97,7% auch als solche erkannt. Die Vokale /a/ und /u/ mit dem intendierten Ausdruck *Wut* der Sprecherin 4 wurden mit 95,3% und 93,2% jeweils sehr gut erkannt. Sonst befanden sich keine *wütenden* Äußerungen unter den 30 am besten erkannten Beispielen.

Es befindet sich nur ein **ängstliches**, von Sprecher 1 gesprochenes /e/ (Bspl. 97) mit 93,2% Erkennungsrate und ein **traurig** geäußertes /o/ der Sprecherin 2 (90,9%) unter den Beispielen 25 – 30, die am besten erkannt wurden.

Anders erkannt als intendiert

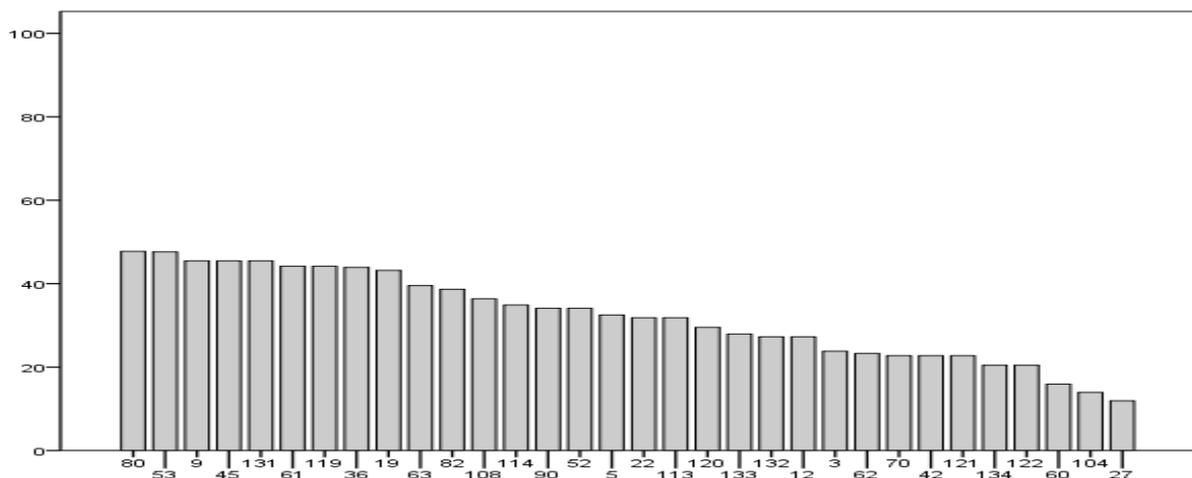


Abb. Die 30 Perzepte, bei denen der intendierte Ausdruck am wenigsten erkannt wurde. (von r. nach l.)

Unter den 30 Perzepten bei denen die angestrebte Emotion am wenigsten erkannt wurde befinden sich nur 3 lautpoetische Wörter. Eines davon (Bsp. 45) wurde von Sprecher 3 mit der Intention **Ekel** ausgesprochen. Zwei weitere Beispiele (119 und 62) wurden vom Sprecher 1 einmal als **Angst** gedacht ausgesprochen und nur von 23,3 % als dieses Gefühl erkannt, und das andere Mal als **Ekel** beabsichtigt und nur von 44,2 % als solches wahrgenommen.

Unter den 20 Beispielen mit der geringsten Trefferquote befindet sich nur ein **lautpoetisches** Wort (Bspl. 62), der Rest sind emotionale Vokale. Fast die Hälfte (9 Beispiele) davon sind die Vokale /u/, /o/ und /a/ mit dem intendierten Gefühl **Ekel** mit einer einzigen Ausnahme des Beispiels 70, in dem /e/ mit der Absicht **Ekel** vom Sprecher 3 gesprochen wurde. Alle anderen Emotionen wurden bei diesem Beispiel mehr wahrgenommen als das beabsichtigte Gefühl von **Ekel**. Die meisten Vpn haben es als **traurig** bezeichnet.

Fehldeutungen

		Bewertung						Total
		Neutral	Wut	Angst	Freude	Trauer	Ekel	
Neutral	Count	859	21	33	58	56	22	1049
	% within Korrekt	81,9%	2,0%	3,1%	5,5%	5,3%	2,1%	100,0%
Wut	Count	59	775	53	84	8	67	1046
	% within Korrekt	5,6%	74,1%	5,1%	8,0%	,8%	6,4%	100,0%
Angst	Count	99	25	554	99	187	79	1043
	% within Korrekt	9,5%	2,4%	53,1%	9,5%	17,9%	7,6%	100,0%
Freude	Count	97	106	32	727	14	72	1048
	% within Korrekt	9,3%	10,1%	3,1%	69,4%	1,3%	6,9%	100,0%
Trauer	Count	171	6	114	32	666	56	1045
	% within Korrekt	16,4%	,6%	10,9%	3,1%	63,7%	5,4%	100,0%
Ekel	Count	83	64	101	23	128	647	1046
	% within Korrekt	7,9%	6,1%	9,7%	2,2%	12,2%	61,9%	100,0%
Total	Count	1368	997	887	1023	1059	943	6277
	% within Korrekt	21,8%	15,9%	14,1%	16,3%	16,9%	15,0%	100,0%

Abb. Fehldeutungen und Verwechslungen

Das größte Missverständnis war bei der Emotion **Angst**, welche häufig mit **Trauer** (17,9%) verwechselt wurde.

Ebenso war die Anzahl der Täuschungen mit den emotionalen Zuständen **Neutral** und **Freude** mit einem Prozentsatz von 9,5% etwas erhöht. Statt **Freude** wurde 106 Mal (10,1%) **Wut** angegeben, weiters wurde die *freudige* oft mit der *neutralen* Sprache (9,3%) verwechselt.

Trauer wurde oftmals irrtümlich als **Neutral** (16,4 %) oder als **Angst** (10,9%) empfunden.

Ekel wurde mehrere Male (12,2%) den Emotionen **Trauer** oder **Angst** (9,7%) zugeschrieben.

Die geringsten Verwechslungen traten bei **Neutral** mit **Wut** (nur 2 %) oder **Ekel** (nur 2,1%), bei **Angst** mit **Wut** (nur 2,4%) und **Ekel** mit **Freude** (nur 2,2%). Kaum Täuschungen waren bei der Emotion **Wut** mit **Trauer** (0,8%), **Trauer** mit **Wut** (0,6%) und mit 1,3% wurde **Freude** 14mal mit **Trauer** verwechselt.

Vpn-Gruppe Zuhause vs. Vpn-Gruppe „Musikzimmer“

Hierbei kam es zu keinen besonderen Unterschied, daher wird in der weiteren Analyse auch keine Differenzierung getroffen.

5.5 Diskussion der Rezipienten Studie

Um eine heterogene Versuchsgruppe zu erreichen, wurden unterschiedliche Berufsgruppen befragt. Doch würde eine gewisse Heterogenität erst ab einer bestimmten Masse erreicht werden. Die TeilnehmerInnen waren alle aus dem beruflichen, künstlerischen, privaten Umfeld (Universität, Freunde, Familie, Beruf, Kunst) der Versuchsleiterin und wurden durch Mundpropaganda oder einschlägigen Email-Verteilern aufgerufen am Experiment teilzunehmen.

Zwar wurde eine große Altersreichweite von 59 Jahren erreicht, aber das Durchschnittsalter ist dem der Versuchsleiterin gleich und könnte noch um einiges ausgeweitet werden.

Leider wurde das angestrebte Mittel beim Geschlecht (also mindestens 50% Männer) nicht ganz erreicht. Man muss dazu sagen, dass sich außer den Studenten der Musikwissenschaft sind, kaum Männer für dieses Thema interessiert haben. Die VI musste männliche Bekannte explizit zur Teilnahme am Experiment bitten. Bei den Frauen war das Interesse hingegen enorm groß.

Die Frage nach der musikalischen sowie stimmlichen Ausbildung war insofern wichtig, als dass man bei den diversen Schulungen mit stimmlichen Ausdrücken unweigerlich in Kontakt kommt und dadurch auch das Hören geschult wird.

Der Ablauf des Hörversuchs erwies sich als zufriedenstellend. Für eine Folge-Studie wäre aber eine Online-Variante des Hörversuchs wegen der größeren Reichweite in andere Länder vorteilhaft. Ebenso wäre dann die Auswertung der Daten nicht so aufwendig wie beim vorliegendem Experiment, wo diese händisch passierte.

Weiters wäre ein Vergleich mit einer Datenerhebung in einer homogenen Gruppe (wie z. B. Psychisch Kranken, Kinder, Pensionisten) interessant.

5.6 Qualitative Erhebung - Auswertung

35 von 44 TeilnehmerInnen beantworteten die Fragen (siehe kap. 5.6) spontan und schriftlich nach Beendigung des Hörversuchs.

Allgemein kann man sagen, dass sehr oft das Merkmal „Klangfarbe“ als Erkennungsmerkmal von emotionalen Zuständen genannt wurde. Weiters wurde von

mehreren Vpn angemerkt, sie hätten die Emotion „Überraschung“ gehört, obwohl diese beim Test nicht wählbar war.

Im folgendem ist eine Auswahl von Antworten zu finden:

Frage 1: *Welche Emotion war deiner/ihrer Meinung nach am besten zu erkennen? Woran lag das?*

Bei diesem Punkt sind am meisten die Emotionen *Ekel* und *Wut* angeführt worden. Als Gründe sind folgende Eigenschaften genannt worden: Tonfall, Lautstärke, Kraft, Tonhöhe, Intonation, Frequenzen, typische Melodie, Klangfarbe etc.

Wut wurde allgemein als kräftiger in der Tongebung (laut) bezeichnet. Eine Person meinte, dass man bei diesen Beispielen „richtig Angst bekommt“.

Ekel sei erkennbar an der „gequetschten Stimme“, innerhalb der Vokale besonders bei /i/ und /e/, an der Melodiekontur, an den gedehnten oder gepressten Lauten etc. Drei Personen schrieben, die vokalen Äußerungen von *Ekel* würden sie an den Brechvorgang erinnern. Eine Vp nannte das „lang gezogene ‚e‘“ und den „glottal Stopp“.

Ebenso wurde *Freude* genannt, da dieses Gefühl eine „typische Tonmelodie“ habe und als einzig positive Emotion wäre Freude leichter von den anderen zu differenzieren. Als weitere spezifische Merkmale wurden bei Freude noch folgende erwähnt: Höhere Frequenzen, „offener und freier Klang“, etc.

Weiters wurde *Neutral* genannt und wie folgt beschrieben: emotionslos, „klingt sauberer“.

Frage 2: *Welche Emotion war am schwierigsten zu erkennen? Warum?*

Unter den Antworten dieser Frage waren die Emotionen *Angst* und *Trauer* am häufigsten zu finden. Als Erklärung wurde u. a. genannt: „schwer abgrenzbar zu anderen Emotionen“, Ähnlichkeiten der beiden Emotionen untereinander und zu anderen wie *Neutral*; „Unterschied zu ‚neutral‘ sehr schwer“, „schwer ohne Gesichtsausdruck zu deuten“, etc.

Weiters hätte *Angst* Ähnlichkeiten mit *Ekel* aufgrund von Sprachmelodie und Timbre. Außerdem sollte doch „ein Zittern in der Stimme“ bei *Angst* zu hören sein.

Darüber hinaus würde *Wut* ähnlich wie *Trauer* oder *Ekel* klingen.

Ekel sei, so eine Vp, „nur in den Vokalen ‚i‘ oder ‚e‘ gut von *Neutral* oder *Angst* zu unterscheiden“.

Eine Person meinte auch, dass die Differenzierung „aller negativen Emotionen“ schwierig gewesen wäre.

Drei weitere Personen gaben das Gefühl *Freude* als problematisch an.

Frage 3: Welche Emotionen sind am schwierigsten zu unterscheiden? Warum?

Die Kombination *Trauer/Neutral* schien am schwierigsten zu differenzieren zu sein und Gründe dafür seien: ähnliche Lautstärke, Tonlage und Intensität; Schwierigkeit Aspekte der Unterscheidung zu finden; beide wären ähnlich ‚tonlos‘; die Äußerungen wären nicht „differenziert genug“;

Weiters war die Unterscheidung von Angst und Trauer problematisch, weil sich diese beiden Emotionen auch ähnlich wären (Tonlage, Intensität, verlangsamte Sprechgeschwindigkeit). Meinungen dazu wurden wie folgt geäußert: „Beides eher unsichere Stimmen“, „Beides sind ‚leere‘ Emotionen“, „werden gesellschaftlich bedingt seltener verbalisiert, daher sind sie ungewohnt und schwer differenzierbar“. Die drei Emotionen *Trauer/Neutral/Angst* bereiteten mehreren Vpn Probleme, da sie keine Anhaltspunkte zur Differenzierung finden konnten. Diese Gefühlsäußerungen seien „ähnlich monoton“ und schwer voneinander abgrenzbar.

Wieder gab eine Vpn „alle negativen Emotionen“ an und meinte, dass sie „oft gleichzeitig auftreten und darum lernen wir nicht sie zu differenzieren, wenn sie mal vereinzelt zu hören sind“.

Wut/Angst wurde von 4 Vpn als schwer differenzierbar genannt, da „sie sich bedingen“ und „ein Aufschrei mit viel Energie kann einerseits *Wut* oder eben auch *Angst* darstellen“.

Angst/Ekel wurde zweimal erwähnt, einmal mit der Begründung, weil sie „dieselbe Tendenz“ hätten und gemeinsam auftreten könnten.

Ebenso wurde *Neutral/Freude* von zwei Personen angegeben, da sie beide „klar und relativ hoch“ oder „weil alles andere negativ ist und diese negative Klang ist irgendwie rauszuhören. Diese beiden sind sich aber ähnlich: klar, offen...“.

Frage 4: Welche Emotionen grenzen sich am deutlichsten voneinander ab?

Wodurch?

Hier wurden am meisten die Gefühle *Freude/Wut* aufgezählt mit den folgenden Unterscheidungsmerkmalen: Tonlage, Lautstärke, Klangfarbe, Melodiekontur, u.a.

Diese Liste zeigt die Emotionspaare, die bei dieser Frage gleich oft genannt wurden, und die Begründung dazu:

Wut und Angst	Intonation, Lautverlauf, Lautstärke („Wut eher laut, Angst eher leise außer bei ‚Angstschrei‘)
Ekel und Wut	bei ersteren wird die Stimme „gequetscht“ und klingt „gedehnt“ oder „schrill“, beim zweiten wird sie „laut“ oder „kurz, schnell, voll und abgehackt“. Ekel klänge nach ‚grauslich‘ Wut hingegen ‚kräftig‘.
Neutral und Wut	unterschiedlicher Tonfall, Tonhöhe anders, Amplitudenunterschied, neutral klingt im Unterschied zu Wut „monoton“
Freude und Ekel	Tonfall, „Bei Freude gibt es eine gewisse energische Leichtigkeit während Ekel eher schwach aber definitiv abwährend wirkt. Das langgezogene ‚e‘ zum Beispiel ist tiefer bei Ekel, lockerer und höher bei Freude.“
Freude und Trauer	Stimm Lage, Tonhöhe, Klangfarbe, Melodieverlauf und „treten selten gemeinsam auf“.

5.7 Diskussion: Qualitative Erhebung

Vorweg sei gesagt, dass fast alle TeilnehmerInnen und jene, die davon erfahren haben, ein großes Interesse am Thema bekundeten.

Die subjektive Einschätzung der Vpn, vor allem der Versuch das Gehörte in Worte zu fassen, war sehr interessant.

Die meisten der Antworten überraschten nicht, da sie im Vorhinein vermutet wurden, wie z.B. dass die Unterscheidung von Neutral, Trauer oder auch Angst schwierig sei oder das Wut leicht erkennbar aufgrund seiner Intensität und Stärke sei. Ebenso lässt die Emotionsauswahl einen großen Interpretationsspielraum, da man alleine die Emotion Angst in unterschiedlichen Nuancierungen aussprechen könnte.

Viele TeilnehmerInnen gaben unmittelbar nach dem Hörversuch auch mündlich an, dass sie auch die Emotion Überraschung gehört hätten. Diese Emotion wurde aber ohnehin im Zuge der Aufnahmen für dieses Experiment genauso wie die anderen Emotionen aufgenommen, also haben die Spr. diesbezüglich eine emotionale Unterscheidung im stimmlichen Ausdruck getroffen. Gerade um der

großen Verwechslungsgefahr zu entgehen und zu aussagekräftigen Ergebnissen zu kommen, hat die VI und ihre Expertenbewerter entschieden, diese sowie die Emotion Scham in diesem Testdurchlauf wegzulassen. Deshalb ist es interessant festzustellen, dass viele der RezipientInnen trotzdem noch die Emotion Überraschung innerhalb der Äußerungen des Hörversuchs vermutet haben. Somit wäre es wichtig die Emotion Überraschung bei einer Folgestudie in die Evaluierung einzufügen. Die Beispiele mit der Intention Überraschung (die es ja schon als Aufnahmen gibt) könnte man dem Test beimengen, während man alle anderen Beispiele gleich lässt.

Interessanter Weise finden sich in den subjektiven Beschreibungen der Vpn ähnliche den emotionalen Lauten zugeschriebenen Attribute, wie sie Trojan (1948) beschreibt (siehe Kap. 2.4, S. 17). Diese Beobachtungen finden sich auch in anderer Literatur (siehe z.B. Paeschke 2003, S. 62–64).

5.8 Akustische Analyse – Zusammenfassung Ergebnisse

Genauere Ausführungen zur Methode, Normen bei der Selektion der Beispiele usw. kann man dem Anhang 2 (S. 63–86) und dem Anhang 3 (S. 87) entnehmen. Dort ist eine Auswahl der akustischen Analyse dargestellt.

Um zu den folgenden Ergebnissen bezüglich F0, F0 Verläufe, F1 & F2 zu kommen wurde mit dem Programm *Praat* gearbeitet. Die Spektralanalyse wurde mittels *Emapson* vollzogen.

Um idiosynkratische Differenzen (z. B. unterschiedliche Grundtonhöhe) zu vermeiden, wurde zuerst die Differenz jedes Sprechers von neutral zur jeweiligen Emotion gemessen und erst dann mit den anderen Sprechern verglichen.

Bei den **lautpoetischen Wörtern** lässt sich folgende Zusammenfassung treffen:

Wut & Zorn:

Generell ist eine Steigerung der F0 zu erkennen. Bei allen Beispielen kann man ähnlich unruhige, nach unten tendierende F0-Verläufe feststellen, die durch ruckartige Spitzen unterbrochen sind. Sie sind den neutralen Verläufen ähnlich aber durch extreme Werte gekennzeichnet. Allgemein lässt sich eine Energiedichte in den unteren Frequenzen bis 1000 Hz. und dann weiters im mittleren bis höheren Bereich

von ca. 1500 bis ungefähr. 3250 Hz. feststellen. Um die 3500 Hz. sinkt die Energie ab. Die Sprechgeschwindigkeit ist beschleunigt.

Angst

Bei dieser Emotion war die niedrigste F0 messbar. Die F0 verläuft gleichmäßig graduell mit einer leichten Tendenz nach unten. Die Artikulation war bei dieser Emotion entweder gesteigert oder vermindert, also ist kein genereller Trend feststellbar.

Freude

Die größte Steigerung der F0 war in der Modulation zum Gefühlszustand *Freude* zu beobachten. Allgemein lässt sich sagen das der F0-Verlauf sehr variantenreich (Höhen und Tiefen) ist und mit einer Ausnahme nach unten ausläuft. Sonst lässt sich keine generelle Tendenz beim F0-Verlauf feststellen. Hohe Energiedichte befindet sich um die 500Hz, dann sinkt die Energie etwas aber verteilt sich fast gleich über mehrere Frequenzen. Man erkennt eine Dichte an Energie in den oberen Frequenzen, jedoch ist die Energie bei ca. 4950Hz wiederum sehr niedrig. Die Sprechgeschwindigkeit ist gesteigert.

Trauer

Vom neutralen zum traurigen Zustand sinkt F0. Der Verlauf von F0 verläuft gleichmäßig ohne extreme Spitzen (Tonhöhenunterschiede) – außer einer Ausnahme – tendenziell nach unten. Bei dieser Emotion gibt es einen Trend zum langsameren Sprechen.

Ekel

Bezüglich der F0 scheiden sich die Geister. Einerseits sinkt sie, bei einem Beispiel ist sie gleichbleibend und beim Beispiel mit der höchsten Erkennungsrate steigt die Grundfrequenz. Der F0-Verlauf zeigt sich variantenreich mit vielen Tonhöhenunterschieden. Energiedichte ist in den unteren Frequenzen am stärksten. Bei dieser Emotion wurde das Wort generell langsamer artikuliert.

Bei den **Vokalen** führte die akustische Analyse zu folgenden Ergebnissen:

/a/

Beim Vokal /a/ kann man von *neutral* zu den Gefühlszuständen *Wut* und *Freude* eine Steigerung der F0 und eine Verschiebung der Formanten F1 u. F2 nach oben beobachten.

Die Spektralanalyse zeigte weiters, dass es bei Freude eine Verdichtung in den oberen Frequenzen gibt und F2 sich nach oben verschiebt. Bei Wut kann in den Formantbereichen noch eine Verstärkung in den unteren Frequenzen um die 300Hz. und in den höheren um die 3000Hz. erkennen.

Bei der Modulation von *Neutral* zu *Trauer* sinkt die F0 und F1 und F2 verschieben sich leicht nach unten.

Die Emotion *Ekel* wurde bei /a/ kaum erkannt, es gibt nur eine Ausnahme: Hier sinkt F0 von *Neutral* zu *Ekel*, F1 sinkt und F2 steigt. F1 befindet sich im für /a/ üblichen Formantbereich und doch erkennt man besonders bei der Spektralanalyse (vgl. Kap., S.), dass auch der F2-Bereich von /e/ erhöht ist, welcher auch als F2 bei dem Vokal /e/ bei der Emotion Ekel derselben Sprecherin auftritt.

/e/

Ekel wurde bei diesem Vokal sehr gut erkannt. Von *neutralem* zu diesem Zustand ist eine Verminderung in der F0 festzustellen. Bei der Spektralanalyse zeigten sich bei Sprecherinnen 2 und 4 ähnliche Werte: Außer den für /e/ üblichen F1 & F2-Bereichen kann man erhöhte Werte in den oberen Frequenzen und bei beiden Sprecherinnen Energieabfall um die 1100Hz beobachten.

Bei der Modulation von *Neutral* zu *Angst* und *Neutral* zu *Freude* ist eine Steigerung in F0 zu erkennen.

Bei der Modulation von *Neutral* zu *Angst*, *Freude* und *Ekel* verschiebt sich F1 nach oben. F2 bei der selbigen Gegenüberstellung geht bei *Angst* nach unten, bei *Freude* etwas nach oben, bei *Ekel* bei einem Bspl. nach unten beim anderen nach oben.

Ein *freudiges* /e/ wurde sehr schlecht erkannt. Bei diesem Beispiel ist eine Verminderung in der F0 feststellbar, wenn man es hört, klingt es sehr dunkel.

/i/

Dieser Vokal mit der Intention *Ekel* ausgesprochen wurde beim Hörversuch sehr gut verstanden. Bei den wesentlichen Beispielen ist bei der Modulation von *Neutral* zu *Ekel* einmal eine Steigerung der F0, zweimal eine Verminderung der F0 zu erkennen. F1 von *Neutral* zu *Ekel* ist bei jedem Sprecher anders: Der Formant sinkt entweder, steigt oder bleibt gleich. F2 sinkt bei allen SprecherInnen.

Niedrige Energiedichte im Bereich von ungefähr 1000 bis ca. 1500 Hz. (ähnlich dem neutralen */i/*) zu verzeichnen. Verstärkt findet sich die Energie aber in den höheren Frequenzen.

F0 bei *Freude* im Unterschied zu *Neutral* ist fast gleichbleibend. F1 und F2 von *Freude* steigen im Unterschied zu *Neutral*. Energiedichte in den höheren Frequenzen.

/o/

Freudige /o/ wurden sehr gut erkannt und es ist jeweils eine Steigerung der F0 von *Neutral* zu *Freude* messbar.

F1 und F2 verschieben sich vom *neutralen* zum *freudigen* Zustand nach oben. Außer in den für */o/* üblichen Formantbereichen kann man bei der Spektralanalyse noch erhöhte Werte in den oberen Frequenzen feststellen.

/u/

Hier sieht man eine Steigerung der F0, f1 und F2 vom *neutralen* zum *freudigen* Gefühl. Die Spektralanalyse zeigt erhöhte Energie in den oberen Frequenzen. Bei *Neutral* zu *Ekel* bleibt F0 und F1 gleich, F2 steigt deutlich.

Diese Erkenntnisse führen zu folgender **Zusammenfassung** bei den Vokalen:

Wut

Wut wurde am besten im Vokal */a/* und auch */e/* erkannt.

Allgemein ist den Vokalen dieser Emotion eine Steigerung der F0, eine Verschiebung von F1/F2 nach oben und erhöhte Energiedichte in den oberen Frequenzen zu erkennen.

Angst

Vokale mit der Emotion Angst geäußert wurden allgemein nicht gut erkannt. Bei den gut erkannten Beispielen kann man folgendes zusammenfassen: F0 steigt. F1 verschiebt sich nach oben, F2 nach unten.

Freude

Freude wurde am besten im Vokal /o/ erkannt. Weiters auch bei einem /u/ und einem /a/.

Kennzeichnend für diese Emotion scheint eine Steigerung der F0, eine Verschiebung von F1 und F2 nach oben und erhöhte Energiedichte in den oberen Frequenzen zu sein.

Trauer

Trauer wurde bei den Vokalen nicht so gut erkannt, am besten noch in /o/.

Festzustellen ist, dass die F0 sinkt, F1 und F2 sich nach unten verschieben.

Ekel

Am eindeutigsten wurde diese Emotion in den Vokalen /i/ und /e/ erkannt.

Bei den Beispielen mit den Vokalen /a/ und /e/ steigt F0, bei den Beispielen mit /i/ gibt es auch Verminderungen der F0. Bei den Formanten F1 und F2 ist keine generelle Tendenz feststellbar.

Es sind ähnliche Tendenzen in den vergleichbaren Punkten bei den Vokalen wie bei den lautpoetischen Wörtern zu entdecken.

5.9 Diskussion: Ergebnisse der Akustischen Analyse und Vergleich mit den Studien

Als statistische Hypothesen werden die Resultate aus der Studie **Banse/Scherer (1996)** herangezogen (siehe kap. 5.2) und mit den Ergebnissen der lautpoetischen Wörter verglichen.

In dem vorliegenden Experiment wurde „*Anger*“ mit „**Wut&Zorn**“ bezeichnet und kann bis auf eine Ausnahme ähnliche Ergebnisse verzeichnen: Steigerung der F0, nach unten tendierende F0-Verläufe. Die Artikulationsrate ist auch hier gesteigert.

Weiters ist eine Energiedichte in den höheren Frequenzen zu erkennen aber dieses Ergebnis ist nicht so entscheidend wie in der Studie Banse/Scherer 1996.

Bei der Emotion **Angst** zeigen sich andere Ergebnisse als bei Banse/Scherer (1996), so wurde bei der eigenen Analyse eine Minderung der F0 festgestellt. Da sich aber bei den Vokalen manchmal andere Ergebnisse zeigen, wird dies wohl eine Frage der Interpretation sein, da man *Angst* sehr unterschiedlich ausdrücken kann. So war auch die Artikulation entweder gesteigert oder gedrosselt. Bei dieser Emotion müsste man genauere Eingrenzungen für Folge-Studien treffen (z.B. Unterscheidung zwischen Sorge, Kummer, Furcht usw.)

Doch die Ergebnisse für das Gefühl **Trauer** sind kongruent mit jenen von Banse/Scherer (1996): Der F0-Verlauf geht tendenziell nach unten und die Sprechgeschwindigkeit verlangsamt sich. Bei dieser Emotion müsste man auch gewisse Eingrenzungen treffen, denn Trauer ist nicht unbedingt dasselbe Gefühl wie Traurigkeit.

Weiters stimmen die Ergebnisse der Emotion **Freude** mit jene von Banse/Scherer 1996 überein: Steigerung der F0, variantenreicher F0-Verlauf, Energiedichte in den höheren Frequenzen und die Sprechgeschwindigkeit ist gesteigert.

Bei der Emotion **Ekel** konnten genauso wie bei Banse/Scherer 1996 keine generellen Tendenzen festgestellt werden außer das der F0-Verlauf variantenreich ist. Bei dem Wort, das 100% Erkennungsrate hatte ist die Artikulation extrem langsam mit längeren Pausen, was aber darin begründet ist, dass Spr. 2 sehr hyperartikulierte spricht, vor allem bei der Emotion *Ekel*. Daraus kann man keine generelle Tendenz ableiten.

Die Annahme zu ähnlichen Ergebnissen wie bei **Banse/Scherer (1996)** hat sich nach Abschluss des Experiments fast ausschließlich bestätigen lassen. Man findet auch Ähnlichkeiten mit den subjektiven Eindrücken der Vpn, dargestellt in der Qualitativen Erhebung (siehe kap. 5.6)

Im Gegensatz zu **Yildirim et al. (2004)** wurden bei diesem Experiment keine generellen Tendenzen bei Trauer festgestellt. Ebenso ist die Sprachgeschwindigkeit bei *Freude* und *Wut* gesteigert. Hier sei aber anzumerken, dass bei Yildirim et al. (2004) nicht *Wut* sondern *Ärger* analysiert wurde. Es ist also anzunehmen, dass man zu anderen Ergebnissen kommt, wenn man die SprecherInnen bittet *Ärger* statt *Wut*

auszudrücken. Daher wäre eine Unterscheidung der Emotionen in Subgruppen bei einem Folgeexperiment (siehe Banse/Scherer 1996) von Vorteil.

Das Sprachmaterial bei Yildirim *et al.* (2004) besteht aus ganzen, semantisch sinnvollen Sätzen und die analysierten Vokale sind diesen entnommen worden. Beim vorliegenden Experiment wurden die Vokale isoliert betrachtet und doch konnten bei den hier verwendeten Vokalen ähnliche Tendenzen bei der F0 wie bei Yildirim *et al.* (2004) festgestellt werden (immer im Vergleich zu *Neutral* zu bemessen):

Zumeist ist auch bei dem vorliegenden Experiment eine Steigerung der F0 bei den Emotionen *Ärger/Wut* und *Freude* ersichtlich. In der Modulation von *Neutral* zu *Trauer* wurde eher eine Verminderung der F0 gemessen.

Wie bei Yildirim *et al.* (2004) zeigte sich, dass *Trauer* und *Neutral* ähnliche F0 Durchschnittswerte haben. Doch nicht immer ist dies der Fall bei *Ärger* bzw. *Wut* und *Freude*. Bei den Ergebnissen des vorliegenden Experiments ist die F0 der *freudigen* Vokale den *neutralen* oft näher als den *wütenden* Beispielen.

Hier müsste noch ein Vergleich mit jenen Vokalen gezogen werden, die aus den lautpoetischen Wörtern entnommen werden. Dann könnten auch Erkenntnisse getroffen werden, wie sich Vokale nach bestimmten Konsonanten verhalten (z.B. nach einem plosiven Laut) und man könnte erforschen, wie sinnvoll manch Stimtraining sind, die ja genaue Vorgaben geben, wie Vokale nach bestimmten Konsonanten ausgesprochen werden (siehe u.a. Balsler-Eberle 1992).

Es können ähnliche Tendenzen von F1 und F2 in der Modulation zu emotionalen Zuständen beobachtet werden, wie es auch Yildirim *et al.* (2004) in ihren Grafiken zeigen (siehe Kap. 5.2).

Die Analyse der Vokale sollte anfangs am ehesten dazu dienen Vokale Emotionen zuordnen zu können. Somit wurden eigentlich nicht wirklich bedeutsame Ergebnisse bei den Vokalen erwartet (da es sich um so kurze Lautäußerungen handelt) und doch konnte die Isolation der Vokale hier aber den Blick auf einige andere Erkenntnisse frei machen:

Bei den Vokalen /i/ und /e/ zeigte sich, dass diese am ehesten Assoziationen mit der Emotion *Ekel* hervorrufen. Das /a/ mit der Emotion Ekel von Spr. 2 ausgesprochen, welches zu 100% erkannt wurde, hat bei genauem Hinhören Tendenzen zum /e/. Die akustische Analyse konnte dies bestätigen. Nur ein /e/ des Spr. 3 wurde kaum als

Ekel erkannt. Interessanterweise hat diese Aufnahme ein anderes Schallbild als die charakteristischen Beispiele für *Ekel*.

Ebenso wie es das Ergebnis bei Yildirim *et al.* (2004) zeigt, erkennt das vorliegende Experiment an, dass der emotionale Effekt auf die Vokale (und dadurch auch die Lage der Formanten) auch mit deren Artikulationsposition und Zungenstellung zu tun hat: Bei diesem Experiment zeigte sich, dass sich *Wut* und *Freude* aber auch *Trauer* eher in den hinteren, offenen Vokalen wie /u/, /o/ und /a/ zu erkennen gibt.

Ekel zeigte sich am besten in den vorderen und geschlossenen Vokalen /i/ und /e/. Ebenso waren einige *freudige* /i/ gut erkennbar, das könnte mit der erhöhten Frequenz zu tun haben, die ein Merkmal für *Freude* aber auch dem Vokal /i/ sehr ähnlich ist.

Manche Formen der *Angst* und *Wut* scheinen sich auch in der Lautäußerung /e/ verwirklichen zu können.

Allgemein kann man sagen, dass die emotionalen Äußerungen auch schlechter differenzierbar sind, wenn sie sich in vielen Parametern (wie F0, F1, F2, usw.) ähneln. Das zeigt die subjektive Betrachtung und bestätigt die objektive akustische Analyse.

Bei den **SprecherInnen** konnte festgestellt werden, dass die beiden professionellen SprecherInnen (Sprecher 3, Sprecherin 4) die beste Erkennungsrate bei den *neutralen* Äußerungen haben, was auf eine besonders akkurate und ausgebildete Sprechweise hindeutet. Die hohe Erkennungsrate bei der semi-professionellen Sprecherin 2 bei vielen emotionalen Ausdrücken ist eher auf die hyperartikulierte Sprechweise hinzuführen, was sich auch darin zeigt, dass diese Spr. bei manchen Äußerungen fast doppelt so langsam spricht wie die beiden Profi-Sprecher. Diese Sprechweise ist wohl nicht als natürlich einzustufen und müsste bei erneuten Aufnahmen kontrolliert werden.

Bei einem Folgeexperiment sollten alle SprecherInnen unter denselben Bedingungen in einem Tonstudio aufgenommen werden. Danach sollte ein Vergleich mit den Beispielen des vorliegenden Hörversuchs gezogen werden.

Weiters wäre es interessant zu sehen, wie sich die hier schon analysierten Beispiele in Kombination mit den Emotionen *Scham* und *Überraschung* verhalten würden, die schon im Zuge der Aufnahmen für dieses Experiment genauso wie die anderen

Emotionen aufgezeichnet wurden. Wie die Auswertung der qualitativen Erhebung (vgl. kap. 5.7) zeigte, hörten einige Vpn *Überraschung* in manchen der Beispielen. Dieses Gefühl scheint genauso wenig differenzierbar wie *Scham*. Auch bei der Studie von Banse/Scherer (siehe 1996, S. 622) waren *Scham* mit einer Erkennungsrate von 22% und Ekel (15%) die Emotionen, die am schwierigsten zu erkennen waren und man meint, dass dies wohl eher am generellen Schallbild der Emotion und nicht beim Ausdruck des Schauspielers liegt (siehe Banse/Scherer 1996, S. 12).

Es wäre auch interessant, wenn die Erwartungshaltung weggenommen wird (siehe Huron 2006) und z. B. keine Emotionen angegeben sind.

Da es universal geltende emotionale Gesichtsausdrücke gibt (vgl. Ekman 1988) und damit schon bei der Datenaquirierung der emotionale Ausdruck besser bestimmt werden kann, wäre die zusätzliche Nutzung von Gesichtsausdrucksaufnahme für multimodale Emotionserkennung von Vorteil (siehe Lee 2004, S. 2).

Messungen der Gesichtsbewegungen mit Hilfe des *Facial Action Coding Systems* (FACS) sind erprobte Hilfsmittel bei der Emotionserkennung (vgl. Ekman 1988, S. 181–224).

Ein weiteres Hilfsmittel wäre die Verwendung von EmuJoy (siehe Nagel *et al.* 2007) bei den Sprachaufnahmen wie bei den Rezipienten, die die Aufnahmen hören und beurteilen, eine zusätzliche Verbesserung in der Klassifizierung und Beurteilung von Emotionen bewirken.

6. Conclusio

Der Klang der Stimme ist etwas sehr individuelles und persönliches. So kann man daraus auch „psychische und adaptive Prozesse“ (Stadler Elmer 2005, S. 126) ableiten. Da der Mensch dazu neigt sich an seine Umgebung anzupassen und adaptiert dabei kulturelle Normen (vgl. Stadler Elmer 2008, S. 145). Diesen Aspekt muss man bei Erstellung von allgemeinen Grundsätzen immer bedenken.

Man kann aber Kausalitäten und Übereinstimmungen in den Beobachtungen z.B. bei Trojan (1948), in dem subjektiven Eindruck wie auch in den akustischen Analysen erkennen.

Es wäre weiters sehr interessant zu noch präziseren universalen Ergebnisse zu kommen, also zu ergründen wo sehr vokale Veränderungen, die von Attitüden und

Emotionen herrühren, kulturell oder sprachlich variieren. Dazu wäre eine großangelegte interkulturelle Erforschung der emotionalen Inferenzen der Stimme und Sprache, die encoders und decoders von verschiedenen Sprachen und Kulturen gut, wie schon Scherer in seinem Conclusio (2000, S. 4) fordert, bei dem der Versuch alle verschiedenen Sprachstämme zu inkludieren (Tonsprachen in Afrika und Asien...) unternommen wird.

Eine Schwierigkeit beim Erkennen emotionaler Lautäußerungen liegt schon in der Definition von Emotion, im Verstehen von Emotion in den Gehirnabläufen von Menschen und fehlende Einigkeit darüber bei Wissenschaftlern. Neue Errungenschaften in den vielen wissenschaftlichen Disziplinen, die mit Emotion zu tun haben, führen auch Steigerungen im Bereich der Emotionserkennung in der Stimme.

Die Erkennung von Emotion in der Stimme ist ein spannendes wie auch sehr zukunftssträchtiges Gebiet der Forschung, da sie eine Verbesserung in der Interaktion Mensch-Maschine darstellt, aber auch wichtige Erkenntnisse für Entwicklungen in der Phonetik, im Fremdsprachenunterricht und für menschliche Dialoge selbst liefert. Es ist anzunehmen, dass es bald nützliche Algorithmen für die automatische Emotionserkennung geben wird.

Die Stimme ist eindeutig ein Übermittler von Stimmungen und der Tonfall ein wichtiger Bestandteil in der menschlichen Kommunikation. In diesem Sinne endet diese Arbeit mit den Worten Trojans:

„Der Ausdruck der Lust schafft die günstigsten Bedingungen für die Entwicklung eines schönen Stimmklanges“

(Trojan 1955, S. 11)

7. Glossar

Diphthong: Die Definition nach Pétursson/Neppert (1991, S. 106) „ein vokalisches Element [...], das aus zwei kontinuierlich ineinander übergehenden Klangfarben besteht und einen Silbenkern bildet.“

HMM: HMM = Hidden Markov Model. Modulation von Sprachcharakteristika auf Basis von parametrischen statistischen Methoden (siehe Schenk/Rigoll S. 126–140)

suprasegmental: „Prosodie umfasst alle suprasegmentalen Merkmale gesprochener Sprache, also alle diejenigen Merkmale, die sich über mehr als ein Segment bzw. einen Laut erstrecken“. (Paeschke 2003, S. 24)

8. Literaturverzeichnis

Akustische Poesie: <http://acoustic-poetry.net/> [04.01.2013]

Ang, Jeremy; *et al.* „Prosody-based automatic detection of annoyance and frustration in human-computer dialog“. University of California 2002.

Banse, Rainer; Scherer, Klaus R. „Acoustic Profiles in Vocal Emotion Expression“. In: *Journal of Personality and Social Psychology*, Vol. 70, No. 3 (1996), S. 614–636.

Birbaumer, Niels; Schmidt, Robert. *Biologische Psychologie*. (Orig. 1945). Heidelberg 2010.

Bruhn, Herbert. „Singen und Erkennen von Melodien“. In: Bruhn, H.; Oerter, R.; Rösing, H., *Musikpsychologie. Ein Handbuch*. Reinbek: Rowohlt 1994, S. 283-290.

Darwin, Charles. *The Expression of Emotions in Man and Animals*. (Orig. London 1872). Übers. von J. Victor Carus. *Der Ausdruck der Gemüthsbewegungen bei dem Menschen und den Thieren*. Stuttgart 1874.

Descartes, René. *Die Leidenschaften der Seele*. (Orig. 1649). Hrsg. u. übers. von Klaus Hammacher. Hamburg 1996.

De Vree, Paul. „Toute predication“. In *zimprovisaties*. Antwerpen 1968.

Dornes, M. Gedanken zur frühen Entwicklung und ihre Bedeutung für die Neurosenpsychologie. In: *Forum der Psychoanalyse* 11(1995). S. 27.–49.

The Nature of Emotion. Fundamental Questions: hrsg. von Paul Ekman und Richard Davidson. New York 1994.

Ekman, Paul. *Gesichtsdruck und Gefühl: 20 Jahre Forschung von Paul Ekman*. Hrsg. u. übers. von Maria von Salisch. (Orig. 1934). Paderborn 1988.

Darwin, Charles. *On the origin of species. By Means of Natural Selection*. Orig. (1859). Hrsg. von Jim Endersby. London 2009.

Frankl, Viktor E. *Logotherapie und Existenzanalyse*. Weinheim 2002.

Georgeva, Ekaterina. *Die Schule für Dichtung in Wien: der Weg der „Mehrkommunikation“*. Dipl. Universität Wien 2009.

Grimm, Michael; Kroschel, Kristian; Mower, Emily; Narayanan, Shrikanth. „Primitives-Based Evaluation and Estimation of Emotions in Speech“. In: *Speech Communication*. (2007).

Hauser, Marc. *The evolution of communication*. Massachusetts 1996.

Hintze, Christian Ide: <http://www.ide7fold.net> [04.01.2013]

Huber, Oswald: *Das psychologische Experiment. Eine Einführung*. Bern 2005.

Huron, David. *Sweet Anticipation. Music and the Psychology of Expectation*. Massachusetts 2006.

Ilie, Gabriela; Thompson, William F. „A Comparison of Acoustic Cues in Music and Speech for Three Dimensions of Affect. In: *Music Perception*. Vol. 23, No. 4 (2006). S. 319–330.

Juslin, Patrik N.; Sloboda, John A. *Handbook of Music and Emotion*. Oxford 2010.

Kleinginna, Paul; Kleinginna, Anne. „A Categorized List of Emotion Definitions, with Suggestions for a Consensual Definition“. In: *Motivation and Emotion*, Vol. 5, No. 4 (1981) S. 345–379.

Kuhl, Julius. *Lehrbuch der Persönlichkeitspsychologie. Motivation, Emotion und Selbststeuerung*. Göttingen 2010.

Lentz, Michael. *Lautpoesie - Lautmusik nach 1945. Eine kritisch-dokumentarische Bestandsaufnahme*. Bd. 1, Wien 2000.

LegéNi: <http://www.sieclelibre.at/index.php?id=6> [04.01.2013]

Lee, Chul Min; Narayanan, Shrikanth. „Emotion Recognition Using a Data-Driven Fuzzy Inference System“. 2003. In: <http://sail.usc.edu>. [04.10.2012]

Lee, Chul Min; Narayanan, Shrikanth. „Toward Detecting Emotions in Spoken Dialogs“. In: *IEEE Transactions on speech and audio processing*, Vol. 13, No. 2 (2005). S. 293–303.

Lee, Chul Min; et al. „Emotion Recognition based on Phoneme Classes“. (2004). In: <http://sail.usc.edu> [01.10.2012]

Melaine, Dolli. *How to take a healthy breath*. Grass Valley, CA 2012 .

Melaine, Dolli. <http://www.voiceopening.net/> [04.01.2013]

Meyer, Jürgen. *Akustik und musikalische Aufführungspraxis*. 5. Auflage. Bergkirchen 2004.

Nagel, Frederik; Kopiez, Reinhard; Grewe, Oliver; Altenmüller, Eckart. „EMuJoy: Software for continuous measurement of perceived emotions in music“. In: *Behavior Research Methods*. Vol. 39 (2), 2007, S. 283–290.

Niedenthal, Paul; et al. „Embodying Emotion“. In: *Science*, Vol. 316 (2007). S. 1002–1005.

Nebert, Ulrich. „Tonhöhe und Sprechstimme – Unterschiede in Mutter- und Fremdsprache“. In: *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 2007.

Oe1 <http://oe1.orf.at/programm/324705> [04.01.2013]

Paeschke, Astrid. *Prosodische Analyse emotionaler Sprechweise*. hrsg. von W. Sendlmeier, Bd. 1, Berlin 2003 (Mündliche Kommunikation).

Parkinson's Voice Initiative: <http://www.parkinsonsvoice.org/> [2013-01-20]

Pétursson, Magnús; Neppert, Joachim. *Elementarbuch der Phonetik*. Hamburg 1991.

Picard, R. W. Affective computing. In: *M.I.T. Media Laboratory Perceptual Computing Section Technical Report*. No. 321 (1995). S. 1–16.

Plutchik, Robert. The Nature of Emotions. In: *American Scientist*, Vol. 89 (2001). S. 344–350)

Plutchik, Robert. *The Psychology and Biology of Emotion*. New York 1994.

Peters, Ralf. „Die schöne und die ganze Stimme – zur ästhetischen Grundlegung der ‚extended voice‘-Bewegung.“, in: Geissner, H.: *Das Phänomen Stimme in Kunst, Wissenschaft, Wirtschaft. 4. Stuttgarter Stimmtage 2002*. S. 45 – 46.

Peterson, G. E.; Barney, H. L. „Control methods used in the study of the identification of vowels“. In: *Journal of the Acoustical Society of America*, Vol. 24, (1952). 175-184.

Rogers, P. L., Scherer, K. R., and Rosenthal, R. „Content-filtering human speech“. In: *Behavioral Research Methods and Instrumentation*, Vol. 3 (1971), S. 16–18.

Reid, Cornelius L. *Funktionale Stimmentwicklung. Zweck und Bewegungsablauf von Stimmübungen*. New York 1988.

Reisenzein, Rainer. „Worum geht es in der Debatte um die Basisemotionen?“. In: *Kognitive und motivationale Aspekte der Motivation*, hrsg. von F. Försterling, J. Stiensmeier-Pelster u. L.-M. Silny. Göttingen 2000. S. 205–237.

Schenk, Jachim; Rigoll, Gerhard. *Mensch-Maschine-Kommunikation. Grundlagen von sprach- und bildbasierten Benutzerschnittstellen*. Heidelberg 2010.

Scherer, Klaus R. „Randomized-splicing: A note on a simple technique for masking speech content“. In: *Journal of Experimental Research in Personality*, Vol. 5, (1971), S. 155–159.

Scherer, Klaus R. „Acoustic concomitants of emotional dimensions: Judging affect from synthesized tone sequences“. In: *Association Meeting*. Boston 1972, S. 105–111.

Scherer, Klaus. „A Cross-Cultural Investigation of Emotion Inferences from Voice and Speech: Implications for Speech Technology“. University of Geneva. 2000.

Schule für Dichtung in Wien: www.sfd.at [04.01.2013]

Schule für Dichtung/Gerhard Rühm: <http://sfd.at/lehren-und-lernen/faculty/gerhard-ruhm> [30.01.2013]

Sedlmeier, Peter; Renkewitz, Frank. *Forschungsmethoden und Statistik in der Psychologie*. München 2008.

Siegwart, Hervine; Scherer, Klaus R. „Acoustic Concomitants of Emotional Expression in Operatic Singing: The Case of Lucia in ‚Ardi gli incensi‘. In: *Journal of Voice* Vol. 9, Nr. 3, S. 249–260.

Stadler Elmer, Stefanie. „Entwicklung des Singens“. In: Stoffer, T. & Oerter, R.: *Allgemeine Musikpsychologie, Enzyklopädie der Psychologie D*, VII, 1. Göttingen 2005, S. 123-152.

Sundberg, Johan. *Die Wissenschaft der Singstimme*. (Orig. 1987). Bonn 1997

Tiedemann, Jens León. *Die intersubjektive Natur der Scham*. Diss. Berlin 2007.

Trojan, Felix. *Die Ausbildung der Sprechstimme. Ratschläge für alle, die beruflich viel sprechen müssen*. Orig. 1895. Wien 1955.

Trojan, Felix, *Der Ausdruck von Stimme und Sprache. Eine phonetische Lautstilistik*. Wien 1948.

Vaitl, Dieter. „Blick ins Gehirn: Wie Emotionen entstehen“. *Giessener Universitätsblätter*; vol. 39; S. 17–24. (2006). (<http://geb.uni-giessen.de/geb/volltexte/2006/3693/>)

Quint, Jutta. *Änderung stimmlicher Parameter, ermittelt durch elektroakustische Messverfahren vor und nach phonochirurgisch-mikrolaryngoskopischem Eingriff*. Dipl. Universität Trier 2005.

Darwin, Charles. *Der Ausdruck der Gemütsbewegungen bei den Menschen und den Thieren*. Stuttgart: E. Schweizerbart 1873.

Original: Darwin, Ch. *The Expression of the emotions in man and animals*. London, 1872.

Spencer, Herbert:

Essays, Scientific, Political and Speculative, 1858. *The Origin and Function of Music*, p. 359.

Yildirim, Serdar; *et al.* „An acoustic study of emotions expressed in speech“. (2004) In: <http://sail.usc.edu> [10.10.2012]

Yildirim, Serdar; Lee, Sungbok; Kazemzadeh, Abe; Narayanan, Shrikanth. „An Articulatory Study of Emotional Speech Production“. In: *Interspeech*. (2005). S. 497–500.

III Anhang

Anhang 1: Abstract zur Diplomarbeit

Abstract (deutsch)

Der Emotionale Ausdruck der Stimme

Die vorliegende Diplomarbeit beschäftigt sich mit dem akustischen Ausdruck von Emotion. Um aus den durch eigene künstlerische Aktivitäten gewonnenen Erfahrungen die unbewiesenen, nicht-wissenschaftlichen Annahmen herauszufiltern, in weiterer Folge eine Trennung von Alltagspsychologie und fundierter Wissenschaft zu erkunden, wird in dieser Arbeit versucht, sich dem Phänomen *Emotion in der Stimme* wissenschaftlich anzunähern.

Zentraler Impetus dabei ist ohne der Ablenkung durch den semantischen Inhalt direkt zum Essentiellen der Stimme, dem Klang und ihrer emotionalen Aussagekraft zu gelangen. Kernpunkt dieser vorliegenden Arbeit ist somit der Vergleich der intendierten Expression in Form von suprasegmentaler Sprache mit den tatsächlichen messbaren akustischen Merkmalen.

Die Abhandlung bedient sich folgender, vor allem in der systematischen Musikwissenschaft angewandter Methoden: der Literaturrecherche, einem explorativen Experiment mit Rezipienten-Test, einer qualitativen Erhebung und der akustischen Analyse.

Einerseits ist ein nomothetischer Ansatz enthalten, in dessen Mittelpunkt die Aggregation von Verhalten und Erleben und die Gewinnung von Durchschnitts- und Normalwerten hieraus steht. Dieser Ansatz soll ermöglichen, unter Annahme einer scheinbar abstrakten Idee eine Theorie abzuleiten. Andererseits ist bei den Ergebnissen des Experiments eine probabilistische Denkweise enthalten, da die Resultate nur wahrscheinlich stimmen und den derzeitigen Stand darstellen. Die Ergebnisse der vorangegangenen Studien konnten bestätigt und um neue Erkenntnisse erweitert werden.

Die Druckversion beinhaltet eine Multimediabeilage (CD).

Abstract (english)

This thesis deals with the acoustic expression of emotion. The work will scientifically approach the phenomenon of *emotion in the voice*, in an attempt to extract any unproven, non-scientific assumptions regarding it. To this end, the argument will in part turn to experiences gained through personal artistic activities, and also explore the divide between folk psychology and aural science. Undistracted by the semantic content, this investigation will then go directly to the Essential of the voice, and the sound of its emotional expressiveness.

The central impetus of this study is the comparison of an intended expression in the form of suprasegmental language with actual measurable acoustic features.

This thesis also employs the primary methodologies used in systematic musicology: literature research, an exploratory experiment with recipient-test, a qualitative survey, and an acoustic analysis.

Furthermore, a mono-thematic approach is included at whose core stands the aggregation of behavior and experience, and the extraction of average and normal values. This approach should help to form a sound hypothesis from an otherwise seemingly abstract idea. Lastly, the outcome of the experiments conclude a probable mindset as their validity is presumed and reflect the current state.

The results of the previous studies were confirmed and extended by new findings.

The printed version contains a CD.

Anhang 2: Akustische Analyse – Eine detaillierte Auswahl

Hauptsächlich konzentriert sich diese akustische Analyse auf die wesentlichen Ergebnisse des Hörversuchs, also jene Beispiele bei denen der intendierte Ausdruck auch als dieser verstanden wurde. Doch werden Beispiele mit niedriger Erkennungsrate, sollte dies nötig sein, zum Vergleich herangezogen.

Teil 1 Gesamtergebnis der lautpoetischen Wörter

Dauer (in sec.)

Spr.	neutral	Wut	Angst	Freude	Trauer	Ekel
1	4,60	3,11	3,14	4,21	3,44	3,57
2	4,39	4,34	3,60	4,41	7,31	8,25
3	2,94	2,43	3,13	2,59	4,06	3,21
4	2,69	2,57	3,27	2,58	2,35	2,30

Abb. Dauer der Wörter in Sekunden

Sprecher 1 und Sprecherin 2 haben eine deutlich niedrigere Artikulationsrate und langsamere Sprechgeschwindigkeit als die anderen beiden SprecherInnen.

Wurde das Wort *wütend* ausgesprochen, dann war die Dauer bei den Sprechern 1, 3 und der Sprecherin 4 kürzer als bei *Neutral*. Bei Sprecherin 2 gibt es hier kaum einen Unterschied. Die lautpoetischen Wörter von SprecherIn 2, 3 und 4 *wütend* artikuliert wurden beim Hörversuch auch von 97,7% der Rezipienten als dieser emotionale Ausdruck wahrgenommen. Wogegen sich das *wütende* Wort von Sprecher 1 zwar nicht unter den 30 am besten erkannten Beispielen befindet, doch aber eine Erkennungsrate von 69,8% hat.

Bei *Freude* steigt die Sprechgeschwindigkeit bei den Spr. 1, 3 und 4. Bei Sprecherin 4 ist diese Steigerung nur minimal.

Die Sprechgeschwindigkeit bei Angst wurde bei Sprecher 1 und Sprecherin 2 im Unterschied zu *Neutral* langsamer, bei Sprecher 3 und Sprecherin 4 wurde sie beschleunigt. Die lautpoetischen Wörter *ängstlich* ausgesprochen konnten bei allen SprecherInnen keine hohe Erkennungsrate erzielen.

Das *traurige* Wort wurde von Sprecherin 4 und Sprecher 1 langsamer als das *neutrale* Wort gesprochen. Sprecher 3 verlangsamte sich beim *traurigen* um mehr als 1 Sekunde als beim *neutralen* Sprechen. Bei Sprecherin 2 waren es gar mehr als 3 Sekunden langsamer als beim *neutralen* Wort. Auch die traurigen Wörter wurden beim Rezipienten-Test nicht deutlich als der intendierte Ausdruck wahrgenommen

daher befinden sich keine unter den 30 Beispielen mit der besten Erkennungsrate. Das *traurig* artikulierte Wort von Sprecherin 2 hatte dabei mit 86,4 % noch den höchsten Identifikationsgrad.

Ekel wurde von Sprecher 1 und 3 langsamer artikuliert als *Neutral*. Sprecherin 2 brauchte für das *ekelhafte* sogar um fast 4 Sekunden länger als für das *neutrale* Wort. Dieses hatte mit 100% die höchste Trefferquote unter den *ekeligen* Wörtern, von denen sich sonst keines unter den 30 am besten erkannten Beispielen befindet. Sprecherin 4 artikulierte das *ekelhafte* ein wenig schneller als das *neutrale* Wort. Dieses wurde beim Rezipiententest mehrmals mit *Trauer* manchmal mit *Neutral* und 1x mit *Wut* verwechselt.

Ergebnisse F0 mittels Praat

F0 wurde an 26 Stellen des Wortes gemessen und daraus der Mittelwert errechnet. Kommastellen wurden entweder auf- oder (unter 5) abgerundet. Settings sind jeweils adaptiert worden. Die beiden Frauen und Sprecher 1 haben ähnliche, Sprecher 3 etwas tiefere Einstellungen.

Sp	Neutr.	Wut	Angst	Freude	Trauer	Ekel
1	213	279	210	300	182	213
2	241	294	227	364	207	260
3	181	205	141	185	110	103
4	215	236	202	236	172	171

Abb. F0 Ergebnisse in Hertz

Bei allen SprecherInnen kann man bei *Wut* und *Freude* gegenüber *Neutral* eine eindeutige Steigerung in F0 erkennen. Die größte Steigerung ist von *Neutral* zu *Freude* zu erkennen. Beim Sprecher 3 ist aber der Unterschied von *Neutral* zu *Freude* nicht wesentlich. Trotzdem ist das lautpoetische Wort von Sprecher 3 *freudig* ausgesprochen, mit 95,3% erkannt worden. Die Steigerung von *Neutral* zu *Wut* war bei Sprecherin 4 nicht allzu hoch und doch wurde dieses Beispiel mit 97,7% Erkennungsrate auch als solches erkannt.

Vom *neutralen* zum *ängstlichen* bzw. *traurigen* Zustand sinkt die F0 wobei sie bei letzterem am niedrigsten ist. Bei Sprecher 1 ist der Unterschied von *Neutral* zu *Angst* kaum merklich und da dieses Beispiel auch beim Rezipienten-Test (nur 23,3%) kaum erkannt wurde, ist dieses Ergebnis nicht relevant.

Ekel hat im Unterschied zu *Neutral* eine niedrigere Frequenz bei Sprecher 3 und Sprecherin 4. Das Beispiel von Sprecher 3 ist unter den 30 mit der niedrigsten Erkennungsrate, also ist es für die Analyse nicht relevant.

Bei Sprecherin 2 wird eine höhere Frequenz gemessen und ist als *Ekel* von 100% benannt worden.

F0 ist den Zuständen *Neutral* und *Ekel* bei Sprecher 1 gleich geblieben. Es ist auch bei den 30 Beispielen mit der schlechtesten Identifizierbarkeit und ist daher nicht weiter bedeutend.

Grundfrequenzverlauf bei allen lautpoetischen Wörtern

Dieser wurde mit dem Programm Praat und denselben Einstellungen, wie bei der F0 Messung, analysiert

Folgende lautpoetischen Wörter hatten sind beim Hörversuch am besten erkannt worden:

SprecherInnen	Emotion	Erkennungsrate
2	Ekel	100%
2	Freude	97,7%
2	Wut	97,7%
3	Wut	97,7%
4	Wut	97,7%
3	Freude	95,3%
4	Neutral	93,2%
3	Neutral	88,6%
1	Neutral	88,4%
4	Freude	86,4%
2	Trauer	86,4%
2	Neutral	81,8%

Abb. Die Erkennungsrate der lautpoetische Wörter je nach Spr. und Emotion

Die andere Hälfte liegt unter 80% und wird daher nicht weiter analysiert. Doch werden einige dieser Beispiele zum Vergleich herangezogen.

Neutral

Spr. 1 Neutral

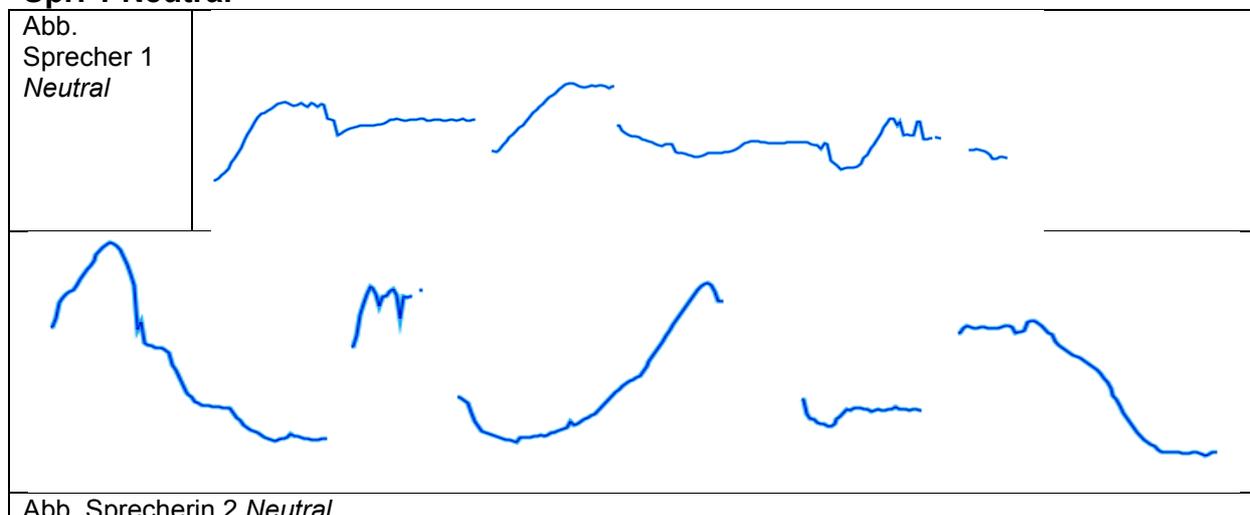




Abb. Sprecher 3 *Neutral*



Abb. Sprecherin 4 *Neutral*

Bei allen Beispielen ist eine Tendenz vom anfänglichen höchsten Punkt im weiteren Verlauf nach unten zu erkennen. Bei den SprecherInnen 2,3 und 4 ist der Schluss auch der tiefste Punkt. Es unterscheidet sich dabei aber das Beispiel der Sprecherin 2 durch einen unruhigeren Verlauf von den anderen.

Wut

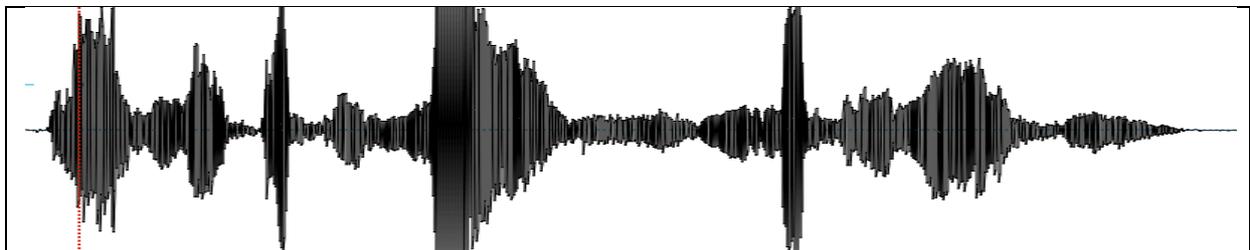


Abb. Sprecherin 2 *Wut*

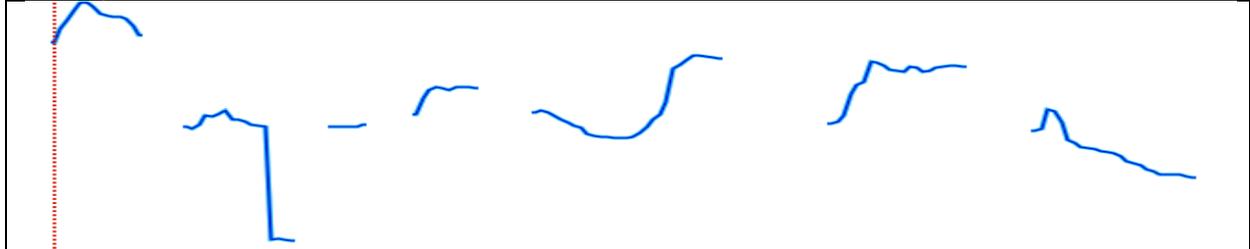
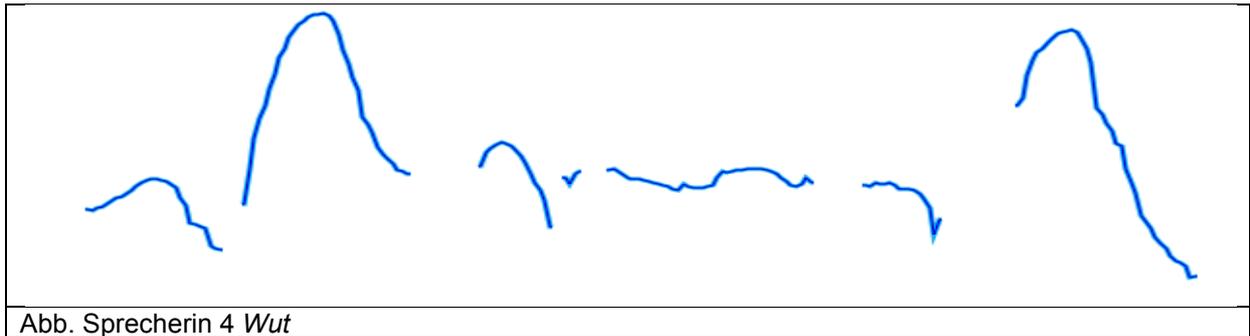


Abb. Sprecher 3 *Wut*

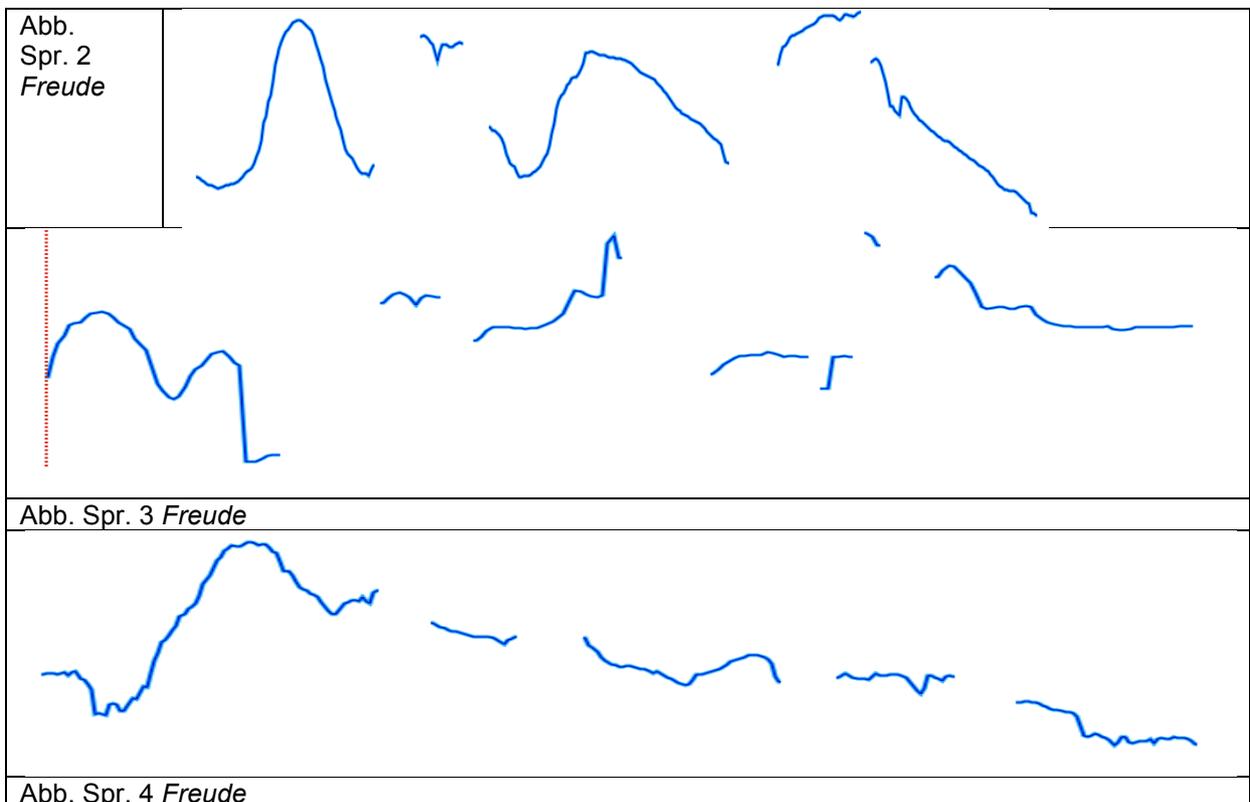


Bei allen diesen Beispielen kann man ähnlich unruhige, ruckartig nach unten tendierende Verläufe beobachten. Es gibt einige markante Spitzen, bei Sprecherin 2 und 4 gibt es zwei sehr hohe Punkte. Der höchste Punkt ist bei Sprecherin 2 bei 431Hz. bei der Sprecherin 4 ist er 379 Hz.

Sprecher 3 setzt wie Sprecherin 2 hoch aber noch ein Stückchen höher als sie ein. Allen gemeinsam ist die abfallende Bewegung am Schluss der bei Sprecherin 4 und 2 auch den tiefsten Punkt und bei Sprecher 3 den zweittiefsten Punkt darstellt.

Bei allen Aufnahmen der wütenden Wörter ist eine leichte Übersteuerung im Ton zu hören.

Freude



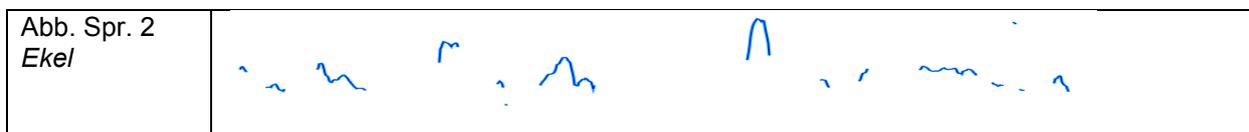
Bei der Emotion *Freude* kann man keinen tendenziellen Grundfrequenzverlauf erkennen. Bei Sprecherin 2 kommt es zu 3 höheren Spitzen, von der letzten läuft die Frequenz stetig bis zum tiefsten Punkt und Schluss hinunter. Beim Sprecher 3 variiert der Verlauf mehrmals und geht tendenziell nach unten, wobei das Ende nicht der tiefste Punkt ist. Bei Sprecherin 4 erlebt man kurz nach Beginn einen Anstieg nach oben bis zum höchsten Punkt bei 366 Hz. Von dort aus verläuft die Frequenz langsam sukzessive bis zum Schluss (tiefster Punkt) nach unten.

Das freudig gesprochene Wort von Sprecher 1 liegt mit 79,1% Erkennungsrate nur knapp hinter den anderen:



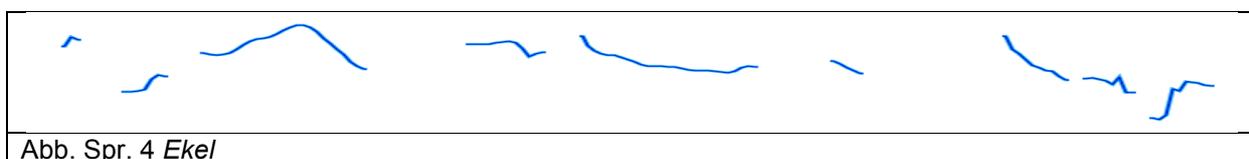
Hier liegt der deutliche Unterschied zu den anderen *freudigen* Beispielen im Anstieg bis zum Schluss, d.h. der zweithöchste Punkt ist auch das Ende und während bei den anderen eine Abwärtsbewegung hin zum Schluss besteht.

Ekel



Das Beispiel *Ekel* von Spr. 2 zeichnet sich durch einen variantenreichen Verlauf mit vielen Tonhöhenunterschieden aus.

Folgendes Beispiel von Sprecherin 4 wurde nur von der Hälfte der HörerInnen als *Ekel* wahrgenommen:



Bei diesem Beispiel kann man ebenfalls einen variantenreichen Verlauf feststellen.

Die Beispiele von den Sprechern 1 und 3 sind kaum als *Ekel* wahrgenommen worden. Bei beiden ist der Verlauf aber weniger variantenreich als bei den oben genannten und gut erkannten Beispielen.

Trauer

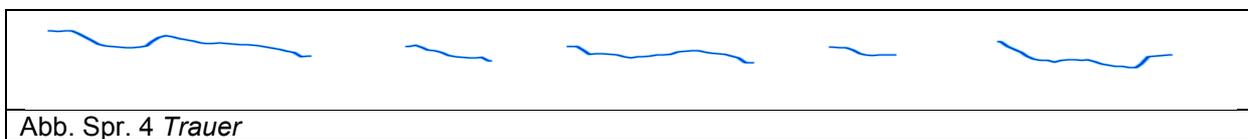
Das *traurige* Wort von Sprecherin 2 gesprochen verläuft gleichmäßig ohne extreme Spitzen vom höchsten Punkt am Anfang bei 222 Hz mit der Tendenz nach unten

gerichtet bis zum Schluss bei 186 Hz, welcher auch der tiefste Punkt ist. Es gibt keine großen Tonhöhenunterschiede.

Das folgende Beispiel von Sprecher 3 wurde von 68,2% als *traurig* wahrgenommen.



Hier ist der Verlauf sehr unruhig mit großen Tonhöhenunterschieden. Auch im Verlauf selbst sind große Unterschiede zu erkennen: Es beginnt bei 142 Hz und endet bei 67 Hz, was eine Verminderung im Verlauf der Frequenz von mehr als die Hälfte ist.



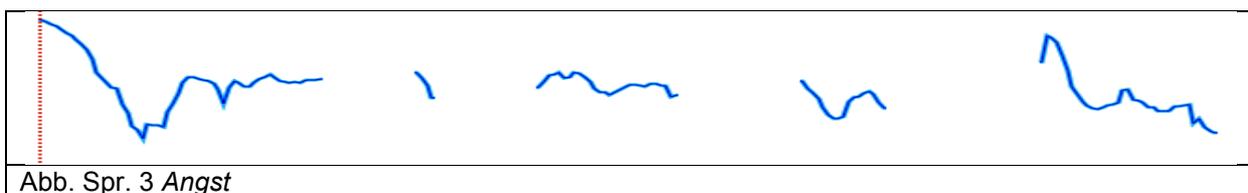
Das Beispiel von Sprecherin 4 wurde von 54,5% als *Trauer* erkannt und verläuft ähnlich gleichmäßig wie jenes von Sprecherin 2. Es beginnt bei 201 Hz und verläuft ohne große Tonhöhenunterschiede tendenziell nach unten bis zum Ende bei 166 Hz. Nur 51,2% nahmen das Beispiel 8 von Sprecher 1 als *Trauer* wahr und dieses hat einen sehr unruhigen Verlauf.

Angst

Unter den Wörtern, die Angst evozieren sollten, wurde das Beispiel von Sprecherin 2 mit 68,2% am besten erkannt.



Der Verlauf hat eher graduelle als ruckartige Veränderungen (außer einem Fall von 214 Hz zum tiefsten Punkt bei 75 Hz). Den gesamten Verlauf betrachtend gibt es aber keinen großen Unterschied vom Beginn bei 254 Hz bis zum Schluss bei 223 Hz. Man kann kaum Vergleiche zum folgendem Beispiel des Sprechers 3 ziehen. Dieses ist von der Hälfte der Rezipienten als Trauer verstanden worden:



Der Beginn stellt den höchsten Punkt bei 190 Hz dar. Der weitere Verlauf ist ohne extreme Spitzen tendenziell nach unten gerichtet. Nur kurz vorm Schluss zeigt sich

ein ruckartiger Anstieg bis 178 Hz, danach verläuft F0 nach unten bis zum Schluss bei 110Hz.

Teil 2 Vokale

F0 und die ersten beiden Formanten der Vokale

Die *neutralen* Werte sind jeweils angegeben, unabhängig davon ob sie herausragende Werte beim Hörtest erzielt haben. Hauptsächlich konzentriert sich die Analyse auf jene Beispiele, die gut von den Rezipienten verstanden worden und schließt manche der Audiofiles zum besseren Vergleich mit ein, die nicht so gut verstanden worden sind.

Bei allen Beispielen wird die F0 Grundfrequenz /a/, F1, F2 als ungefährender Mittelwert (an 10 Stellen gemessen und dann Mittelwert errechnet) dargestellt.

Als Vergleich wurden die gemessenen Formantbereiche von Meyer (2004, S. 33) und Peterson/Barney (1952) unter den eigenen Messungen geschrieben. Die Abkürzungen bei Peterson/Barney (1952) bedeuten folgendes: M = Mann; W = Frau; CH: Kind;

Vokal /a/

F0 Grundfrequenz /a/ ungefährender Mittelwert (an 10 Stellen gemessen und dann Mittelwert errechnet)

Spr.	Neutr.	Wut	Angst	Freude	Trauer	Ekel
1	193			350		321
2	235	284	255	247	203	190
3	144				103	123
4	175	196		277		189

Abb. F0 von Vokal /a/

Von Sprecher 1 wurde kein emotional geäußertes /a/ gut verstanden. Die Vokale, die mit den Emotionen *Freude* und *Ekel* gesprochen wurden, befinden sich unter den 20 mit der niedrigsten Erkennungsrate. Der *neutrale* Zustand wurde von 76,7% als solcher erkannt.

Die Vokale /a/ mit den Intentionen *Freude* und *Ekel* von Sprecherin 2 artikuliert, wurden sehr gut verstanden. Die Vokale /a/, *wütend* und *traurig* kundgetan, hatten auch hohe Erkennungsraten von 86,4% bzw. 86,1%. Das *ängstliche* /a/ wurde bei

dieser Sprecherin nicht gut erkannt. Die *neutrale* Form wurde von 75% erkannt, also von insgesamt 11 Personen anders verstanden.

Das *traurige /a/* und das *ekelige /a/* von Sprecher 3 befinden sich unter den 20 Beispielen, die am wenigsten erkannt wurden. Das */a/* von Sprecher 3 in *neutraler* Befindlichkeit gesprochen hat eine sehr hohe Erkennungsrate von 97,7%, also nur eine Person konnte diesen Zustand nicht als solchen deuten.

Bei Sprecherin 4 sind die */a/-*Beispiele mit den evozierten Emotionen *Wut* und *Freude* sehr gut verstanden worden. Das *neutrale /a/* hatte mit 90,9% ebenfalls eine hohe Erkennungsrate. Das */a/* mit der Intention *Ekel* von Sprecherin 4 geäußert wurde sehr schlecht erkannt.

F1

Spr.	neutral	Wut	Angst	Freude	Trauer	Ekel
1	983			1001		983
2	890	1001	760	1020	760	834
3	834				780	816
4	1001	1149		1001		1038

Abb. F1 von /a/

F2

Spr.	neutral	Wut	Angst	Freude	Trauer	Ekel
1	1400			1590		1557
2	1316	1740	1390	1630	1298	1870
3	1242				1335	1483
4	1446	1520		1480		1502

Abb. F2 von /a/

Die ersten beiden Formanten sind laut Peterson/Barney (1952, S.183):

F1: M: 730; W: 850; CH: 1030 F2: M: 1090 W: 1220 CH: 1370

F1 und F2 sind laut Meyer (2004, S. 33): F1: 1000 F2: 1400 Hz

/e/

F0

Spr.	neutral	Wut	Angst	Freude	Trauer	Ekel
1	160		262	384		
2	234			325		211
3	136			122		132
4	197					177

Abb. F0 von /e/

Unter den am besten erkannten Beispielen des Hörversuches befinden sich die Beispiele *Neutral* und *Angst* von Sprecher 1, *Ekel* und *Freude* von Sprecherin 2,

Neutral und *Ekel* von Sprecherin 4. Das *neutrale* Beispiel von Spr. 2 wurde mit 75% genügend und jenes von Spr. 3 mit 90,9% erkannt.

Unter den 20 Beispielen, die am wenigsten erkannt wurden befinden sich folgende Beispiele: *Freude* von Sprecher 1, *Ekel* und *Freude* von Sprecher 3.

F1

Spr.	neutral	Wut	Angst	Freude	Trauer	Ekel
1	408		686	590		
2	426			500		460
3	297			297		371
4	390			742		400

Abb. F1 von /e/

F2

Spr.	neutral	Wut	Angst	Freude	Trauer	Ekel
1	2130		1828	2080		
2	2520			2580		2225
3	2466			2577		2280
4	2725			2650		2800

Abb. F2 von /e/

F1 und F2 sind bei Peterson/Barney (1952, S.183):

F1: M: 530 W: 610 CH: 690 F2: M:1840 W: 2330 CH: 2610

Die Formantbereiche für /e/ sind laut Meyer (2004, S. 33): F1: 500 F2: 2300 Hz

/i/

F0

Spr.	neutral	Wut	Angst	Freude	Trauer	Ekel
1	196	213				216
2	263		261	260		380
3	166			153		153
4	186					167

Abb. F0 von /i/

Sehr gut erkannt wurden die Beispiele: *Ekel* von Sprecher 1, *Freude* und *Ekel* von Sprecherin 2, *Neutral* und *Ekel* von Sprecher 3, *Ekel* von Sprecherin 4.

Unter den letzten 30 Beispielen, die am wenigsten erkannt wurden befinden sich:

Wut von Sprecher 1, *Angst* von Spr. 2 und *Freude* von Spr. 4.

Die *neutralen* Beispiele sind bei Spr. 3 von 95,5%, bei Spr. 2 von 86,4%, bei Spr. 4 von 81% und bei Spr. 1 von 56,8% der Vpn erkannt worden.

F1

Spr.	neutral	Wut	Angst	Freude	Trauer	Ekel
1	315	440				315
2	334		297	445		334
3	334			297		315
4	330					352

Abb. F1 von /i/

F2

Spr.	neutral	Wut	Angst	Freude	Trauer	Ekel
1	2280	2210				2188
2	2633		2700	2840		2559
3	2800			2800		2670
4	2880					2500

Abb. F2 von /i/

Die Bereiche für F1 und F2 sind bei Peterson/Barney (1952, S.183):

F1: M: 270/390 W: 310/430 CH: 370/530 F2: M:2290/1990 W: 2790/2480 CH: 3200/2730

Die Formantbereiche für den Vokal /i/ sind bei Meyer (2004, S. 33):

F1: 320 F2: 3200 Hz

/o/

F0

Spr.	neutral	Wut	Angst	Freude	Trauer	Ekel
1	174	274				
2	237			388	231	
3	152					135
4	206			267		138

Abb. F0 von /o/

Das *wütende /o/* von Spr. 1 und die *ekeligen /o/* von Spr. 3 und 4 sind unter den 20 Beispielen, die am geringsten erkannt wurden.

Von Spr. 2 befinden sich die Emotionen *Freude* und *Trauer*, von Spr. 4 das Gefühl *Freude* unter den /o/ Ausdrücken, die am besten erkannt wurden.

Die *neutralen /o/* von Spr. 1, 3 und 4 sind allesamt gut erkannt worden (72,7%, 84,1% und 76,7%).

/o/ von Spr. 2 *neutral* geäußert wurde nur von 57,1% der HörerInnen erkannt.

F1

Spr.	neutral	Wut	Angst	Freude	Trauer	Ekel
1	426	630				
2	463			538	482	
3	315					428
4	426			593		463

Abb. F1 von /o/

F2

Spr.	neutral	Wut	Angst	Freude	Trauer	Ekel
1	742	1020				
2	742			983	871	
3	575					630
4	871			983		853

Abb. F2 von /o/

Die Formantbereiche (F1, F2) von /o/ sind laut Peterson/Barney (1952, S.183):

F1: M: 570 W: 590 CH: 680 F2: M:840 W: 920 CH: 1060

F1 und F2 sind bei Meyer (2004, S. 33): F1: 500 F2: 1000 Hz

/u/

F0

Spr.	neutral	Wut	Angst	Freude	Trauer	Ekel
2	236			378		238
3	157		205	186	200	183
4	202					175

Abb. F0 von /u/

Von Spr. 1 ist kein gut erkanntes Beispiel für /u/ zu beschreiben. Bei Spr. 2 wurde *Freude* und *Ekel* sehr gut erkannt. Bei Spr. 3 befinden sich keine Beispiele unter jenen, die gut erkannt wurden. Folgende Beispiele sind unter den 20, die am wenigsten erkannt wurden: *Ekel*, *Angst*, *Trauer* und *Freude*. Ebenso wurde bei Spr. 3 kein /u/ sehr gut erkannt und unter den letzten 20 ist das *ekelige* /u/ zu finden. Das *neutrale* /u/ von Spr. 2 wurde mit 75%, von Spr. 3 mit 68,2% und von Spr. 4 mit 88,6% erkannt.

F1

Spr.	neutral	Wut	Angst	Freude	Trauer	Ekel
2	352			445		352
3	315		390	278	426	389
4	352					352

Abb. F1 von /u/

F2

Spr.	neutral	Wut	Angst	Freude	Trauer	Ekel
2	816			945		1335
3	1170?		1020?	1150?	650	650
4	964					1430 ?

Abb. F2 von /u/

Die Formantbereiche (F1, F2) für /u/ sind laut Peterson/Barney (1952, S.183):

F1: M: 440/300 W: 470/370 CH: 560/430 F2: M:1020/870 W: 1160/950 CH: 1410/1170

F1 und F2 für diesen Vokal sind laut Meyer (2004, S. 33): F1: 320 F2: 800 Hz

Teil 3: Spektralanalyse mittels Emapson

Im folgendem wurden die 25 am besten erkannten Beispiele analysiert. Bei den Ergebnissen handelt es sich über den gesamten Zeitbereich der Lautäußerung gemittelte Werte.

Darunter befinden sich folgende lautpoetische Wörter:

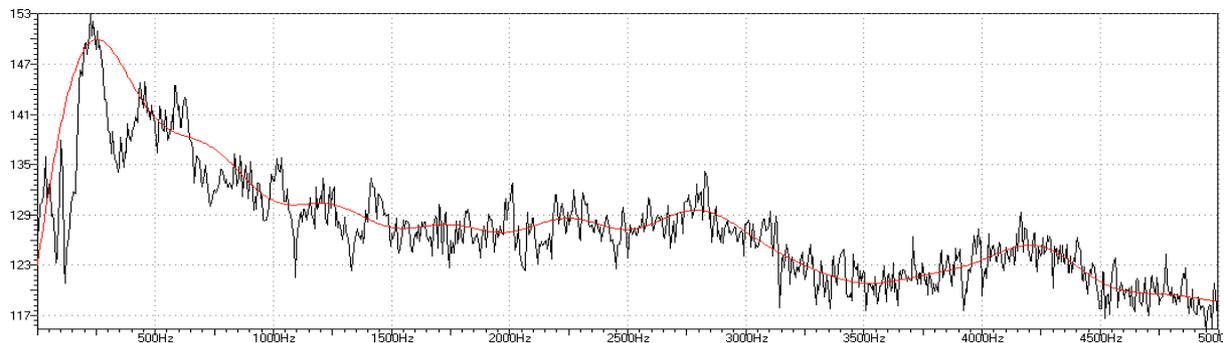


Abb. Spr. 4 neutral

Die größte Spitze ist um die 250 Hz, die weiteren Spitzen bei 450 Hz und 600 Hz zu sehen. Die Energie sinkt bei 3500 Hz ab.



Abb. Spr. 2 *Ekel*

Man sieht eine Verstärkung in den unteren Frequenzen unter 100Hz. Spitzen sind bei 250 Hz, 400 Hz und 500 Hz zu sehen. Erhöhte Bereiche sind von 600–950 Hz, 1050–1150 Hz., 1300–1750 Hz. und bei 1800 bis 1950 Hz. Danach geht der Verlauf nach unten wobei ab ca. 2800 bis 4600 Hz. abermals ein etwas erhöhter Bereich besteht.



Abb. Spr. 2 *Wut*

Hier sieht man eine deutliche Erhöhung zwischen 250 und 750 Hz. Die höchste Spitze liegt bei ca. 700 Hz. Bei 1400 und 1500 Hz. Danach fällt die Energie stark nach unten. Danach steigt die Energie wieder etwas und somit ist der Bereich zwischen 1600 und 3000 Hz. erhöht.

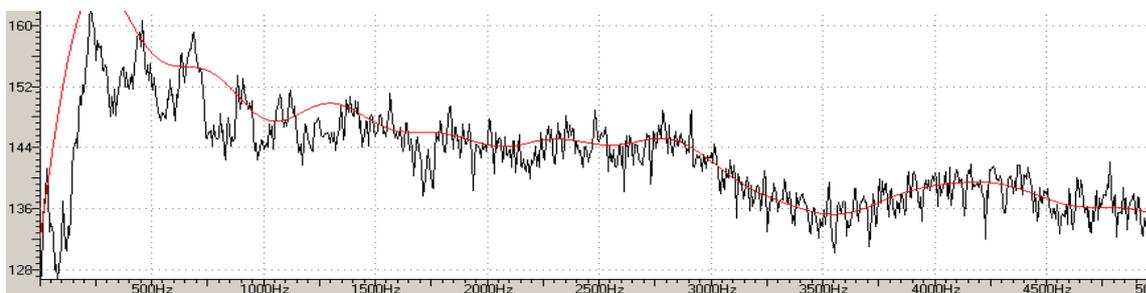


Abb. Spr. 4 *Wut*

Am merklichsten sind die Frequenzen von 250 bis 750 Hz. verstärkt. Dann gibt es weiter eine Erhöhung im Bereich von 900 bis 2900 Hz.

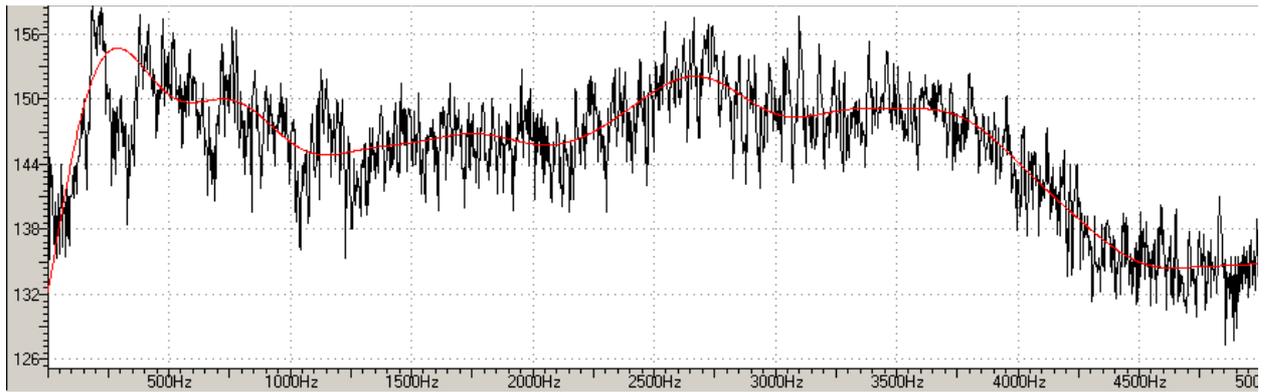


Abb. Spr. 3 *Wut*

Die Spitzenwerte bei diesem Beispiel sind bei 250 bis 750 Hz. Der zweite große Bereich ist von 2500 bis 3500 Hz.

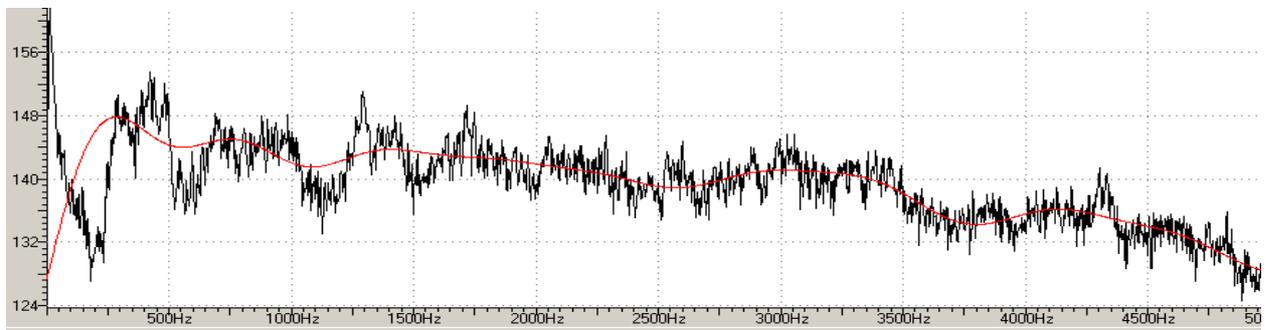


Abb. Spr. 2 *Freude*

Erhöhte Bereiche sind von 300–500 Hz, 650–1000Hz. und 1250–1450 Hz. Weiters ist der Bereich von 2100 bis 3500 Hz etwas erhöht, ab dann beginnt die Energie zu sinken.



Abb. Spr. 3 *Freude*

Die höchste Spitze findet sich bei 66 Hz. Weitere Spitzen liegen bei 250 Hz, 450 Hz, 800 Hz, 1000 Hz, 1300 Hz, 1600 und 1800 Hz. Der Bereich von 3400 bis 3850 Hz ist deutlich erhöht. Weiters gibt es starke Frequenzen im Bereich von 2550 bis 2800 Hz.

Unter den 20 Beispielen, die am **wenigsten erkannt** wurden befindet sich das Beispiel Nr. 62 von Sprecher 1 (23,3% Erkennungsrate):

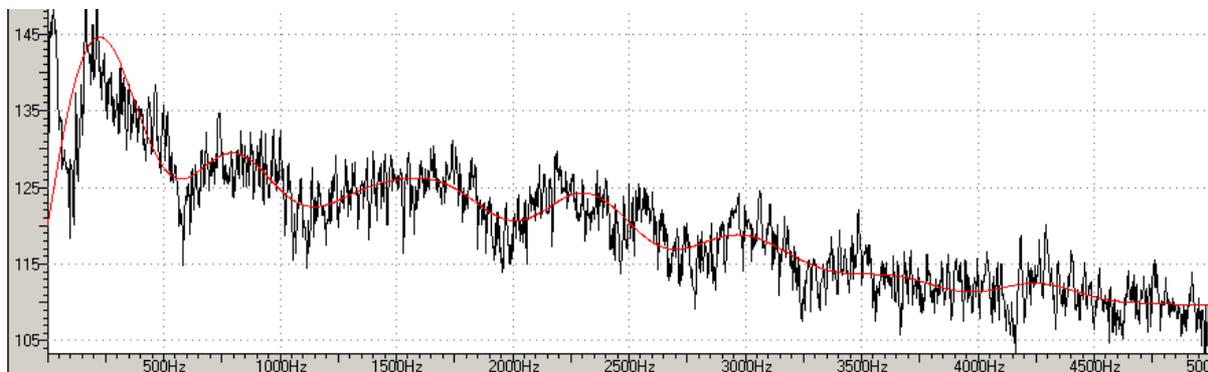


Abb. Spr. 1 *Angst*

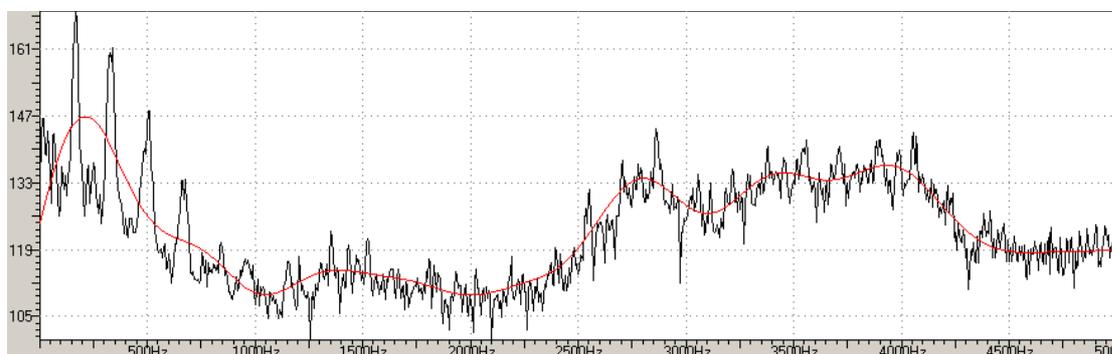
Hier zeigt sich eine deutliche Erhöhung zwischen 220 bis 500 Hz. Der Bereich 750 bis 1000 Hz ist am zweitstärksten, dann der Bereich 1200 –1800 Hz. In Richtung der höheren Frequenzen ist die Tendenz fallend.

Vokale

Spektralanalyse mittels *Emapson*

Folgende Darstellungen zeigen jene akustischen Äußerungen, die unter den 25 am besten erkannten Beispiele sind:

Neutral



Spr. 3 */i/ Neutral*

Spitzenwerte finden sich bei 200 Hz, 350 Hz und 500 Hz. Eine weitere Erhöhung ist bei 600 Hz zu sehen. Dann sinkt die Energie. Ab 2600 bis 4250 Hz verdichtet sich der Energiebereich wieder.

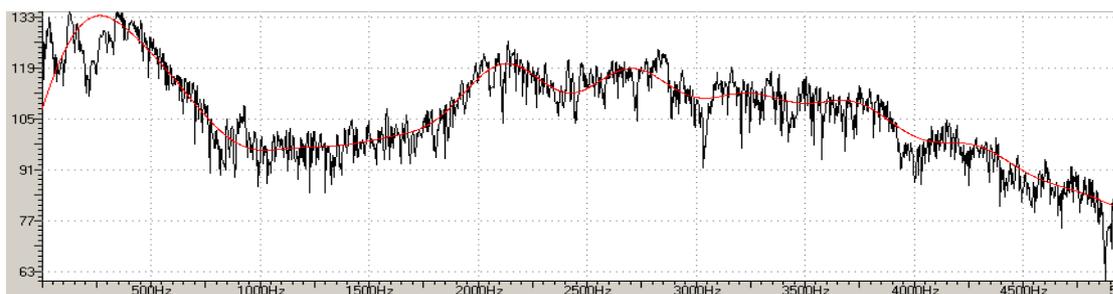


Abb. Spr. 1 /e/ Neutral

Spitzen sind bei 50Hz, 200 Hz, und 400 Hz zu sehen. Nach der letzten Spitze schwindet die Energie sukzessive. Zwischen 2000 bis 3900 Hz findet sich ein weiterer erhöhter Bereich.

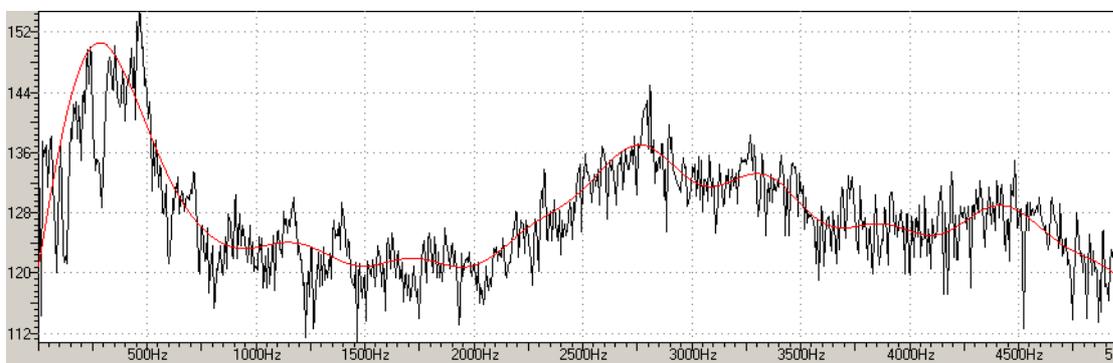


Abb. Spr. 4 /e/ Neutral

Spitzen sind bei 250 Hz, 400 und 500 Hz zu verzeichnen. Danach ist ein Bereich mit weniger Energie. Ab 2500 bis 4500 ist die Energie wieder stärker mit einer Spitze bei 2850 Hz.



Abb. Spr. 3 /a/ Neutral

Die Energie gipfelt in den Spitzen bei 150Hz.und 300 Hz. In den erhöhten Bereichen von 750–950Hz und 1100 bis 1350Hz. finden sich gleich mehrere Spitzen. Danach sinkt die Energie für eine Weile. Zwischen 2750–2900 und 3750–3900 Hz ist sie abermals stärker. Ab 4250 Hz. entschwindet die Energie.

Freude

Es befindet sich nur 1 **Vokal /u/** darunter:

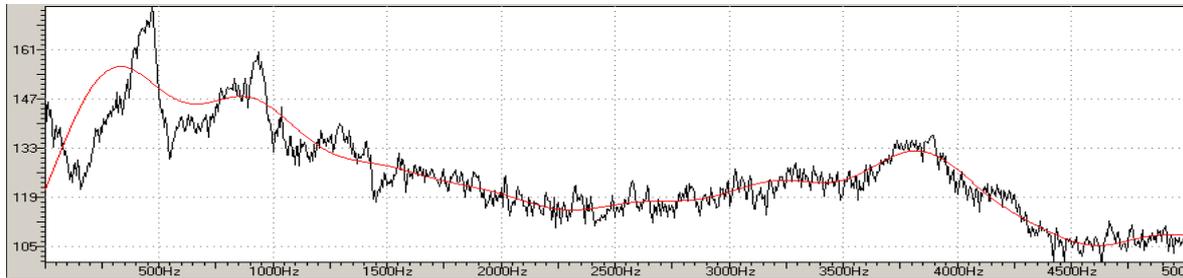


Abb. Spr. 2 *Freude*

Starke Erhöhung bei 445Hz (= F1 von *freudigem /u/* bei Spr. 2). Zweite Spitze sieht man bei ca. 945 Hz (= F2 des *freudigen /u/* dieser Sprecherin). Eine weitere Erhöhung ist in den oberen Frequenzen zwischen 3750 und 3900 Hz.

Weiters wurden die ***freudig* geäußerten Vokale /o/** der Sprecherin 2 und 4 sehr gut erkannt:



Abb. Spr. 2 */o/ Freude*

Die erste Spitze ist bei (538 Hz = F1 des *freudigen /o/* Spr. 2). Weitere Spitzen sind bei 983 Hz (= F2 des *freudigen /o/* Spr. 2) und 1000 Hz (= üblicher Formantbereich von /o/). Weitere Verstärkungen finden sich von 2800–3100 Hz und 4250-4750 Hz.



Abb. Spr. 4 /o/ Freude

Höchste Spitze sieht man bei 700 Hz. Weitere Spitzen finden sich bei 350 Hz und 538 (= F1 des freudigen /o/ dieser Spr.) und bei 900 bis 1000 Hz (üblicher F2 Bereich). Ein weiter erhöhte Bereiche von 100 bis 1400 Hz. Eine nochmalige Erhöhung findet sich um die 3400 Hz.

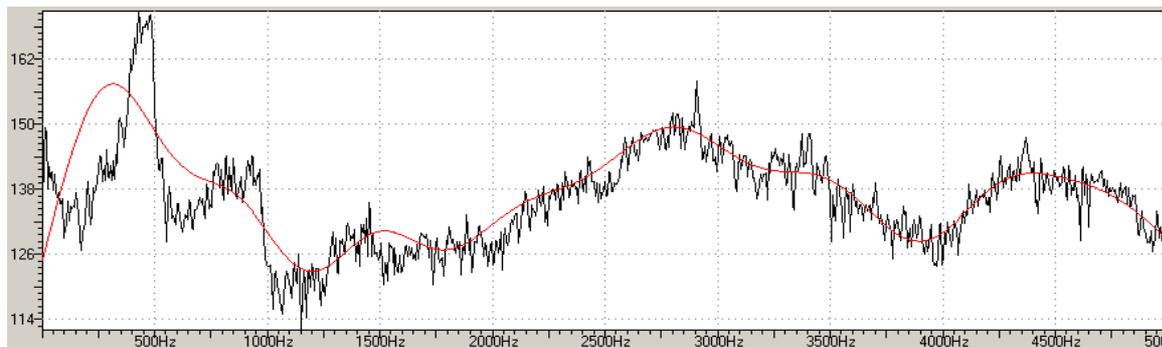


Abb. Spr. 2 /i/ Freude

Höchste Spitze bei 490 bis 500 Hz, das liegt im selben F1-Bereich von /o/ und /e/. Bei dieser Sprecherin wurde beim *freudigen* /i/ als F1-Wert 445 Hz gemessen. Dieser ist um einiges höher als der übliche F1-Bereich von /i/. Eine weitere Erhöhung findet sich in den üblichen F1 und F2-Bereichen von /i/ (320 Hz/3200 Hz). Es ist eine Energiedichte in den höheren Frequenzen zu sehen.



Abb. Spr. 2 /e/ Freude

Bei diesem Beispiel erkennt man eine Energiedichte um die 2400 bis 3250 Hz. Weiters ist der Bereich von 4250 bis 5000 Hz etwas erhöht. Der erste Bereich ist F2 von /e/, *freudig* gesprochen von Sprecherin 2. Die höchste Spitze ist bei 490 Hz, die zweitstärkste Spitze bei 750 Hz. Erstere ist F1 von vom *freudigen* /e/ dieser Sprecherin.

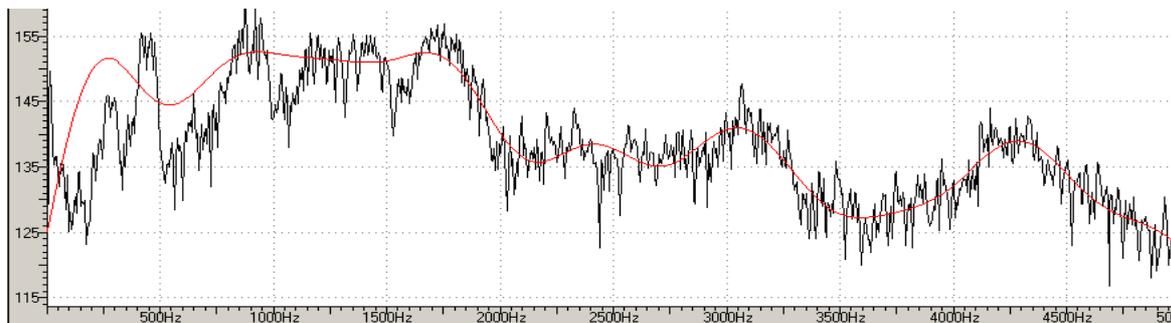


Abb. Spr. 2 /a/ Freude

Die meiste Verstärkung ist zwischen 800-1450 Hz, mit Spitzen bei 450 Hz und 900 Hz. Weiters findet sich eine stärkere Energiedichte um die 1800 Hz, 3100 Hz und in den höheren Frequenzen um die 4250 Hz. Außer den typischen Formantbereichen für /a/ ist also eine Verstärkung in den oberen Frequenzen und eine Erhöhung von F2 zu erkennen.

Wut

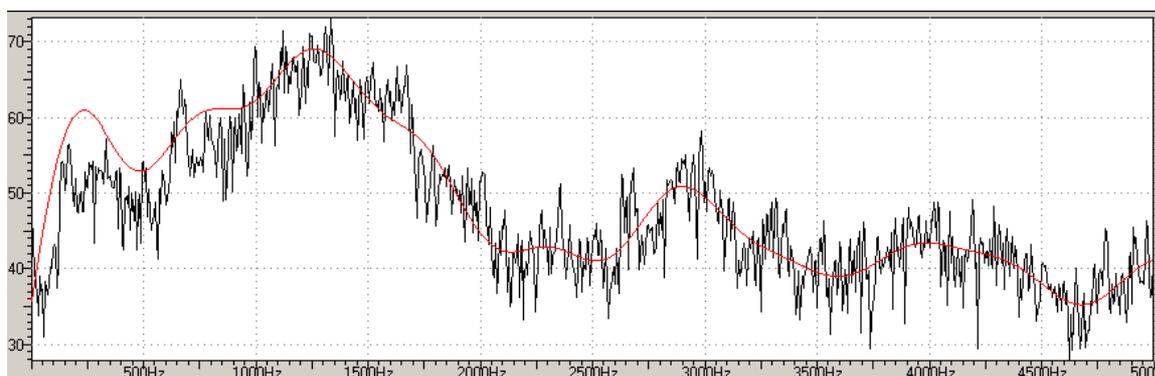


Abb. Sprecherin 4 /a/ Wut

Erhöhte Werte sind vor allem im Bereich zwischen 1000Hz. und 1750Hz zu erkennen. Ebenso ist der Bereich von 200 bis 480Hz erhöht. Niedriger Wert von 1600 bis ca. 2600Hz., dann verstärkt sich Energie bei um die 3000 Hz. wieder.

Angst

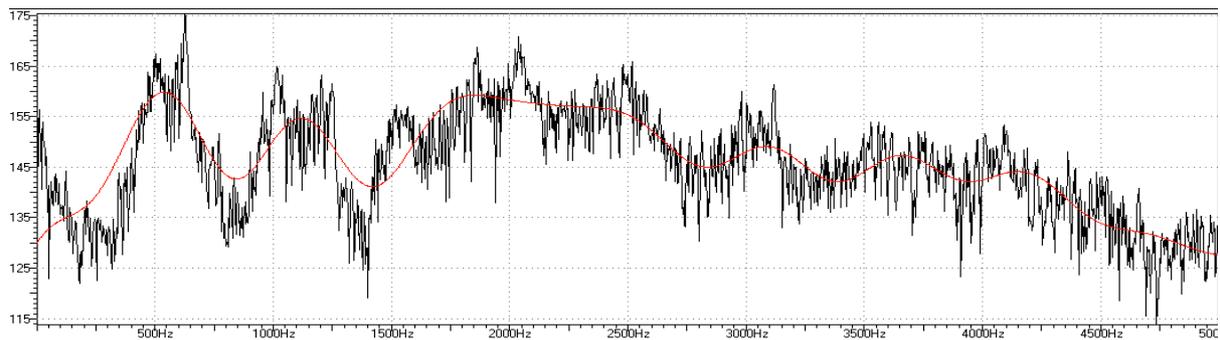


Abb. Spr. 1 /e/ Angst

Hier sieht man mehrere Spitzenwerte und erhöhte Bereiche (bei um die 500 bis 600Hz; 1000–1250Hz und bei um die 3000Hz). Längster Bereich verdichteter Energie von 1500Hz.–2500Hz. Erst ab 4750Hz. sinkt der Energiewert.

Ekel

Vokal /a/

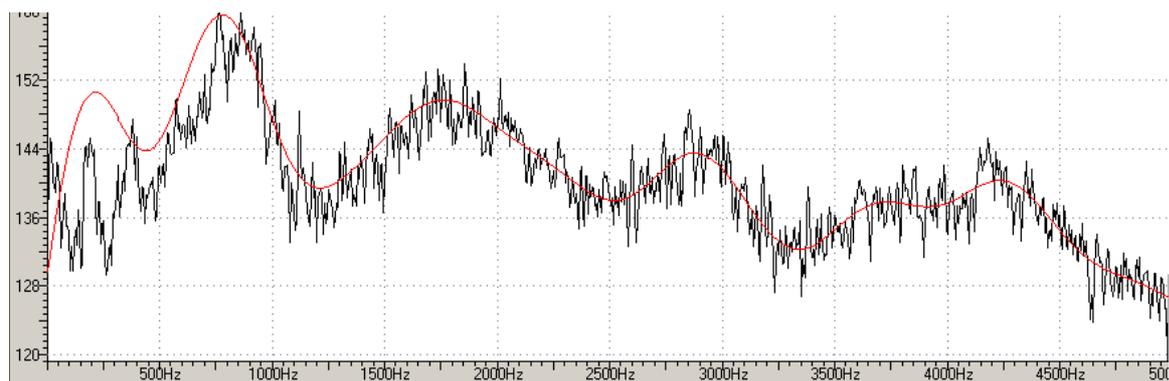


Abb. Sprecherin 2 /a/ Ekel

Dieses Beispiel hatte 100 % Erkennungsrate. Die größte Verstärkung ist bei 750–950 Hz. Weitere Bereiche mit erhöhter Energie sind: 1600–2000 Hz, 2250–3000 Hz und bei um die 4250 Hz. Einerseits sind die Bereiche F1 & F2 von /a/, andererseits aber auch jener von F2 bei Vokal /e/ erhöht. Es ist auch genau jener F2-Bereich von Spr. 2 erhöht, den sie als F2 bei ihrem als Ekel geäußerten Vokal /e/ hat. Wenn man das Beispiel hört, klingt das /a/ im zweiten Teil auch eher ähnlich einem /e/.

Die mit der Intention Ekel geäußerten /i/ der Sprecher 1, 3 und 4 sind in folgenden Beispielen dargestellt:

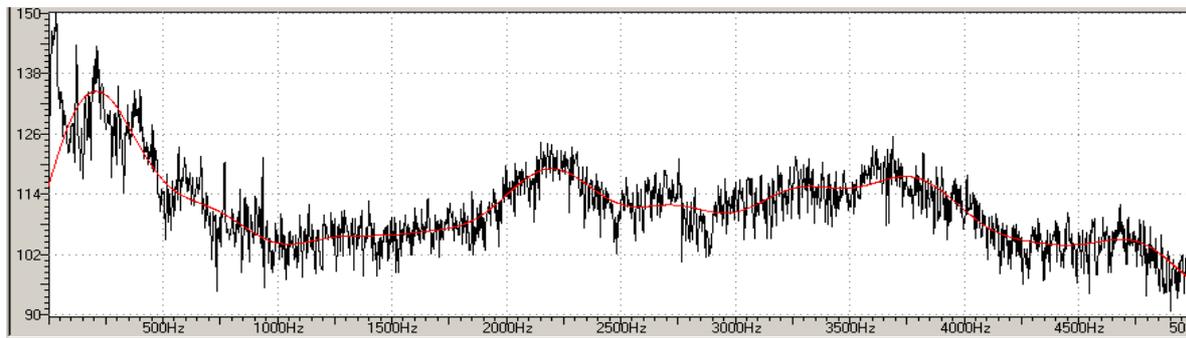


Abb. Spr. /i/ 1 *Ekel*

Die höchsten Spitzen sind bei 250Hz, dann bei 400 Hz. Ein erhöhter Bereich ist von 550 bis 6500 Hz. Dann sind weitere Energiedichten von 2000 bis 2400 Hz und von 3000 bis 4000 Hz.



Abb. Spr. 4 /i/ *Ekel*

Hier sind Spitzenwerte bei 250, bei 300 und 400 Hz zu sehen. Eine weitere Spitze befindet sich bei um die 600 Hz, dann wieder bei ca. 800 Hz. Der Bereich von 2300 bis 2900 ist verstärkt.

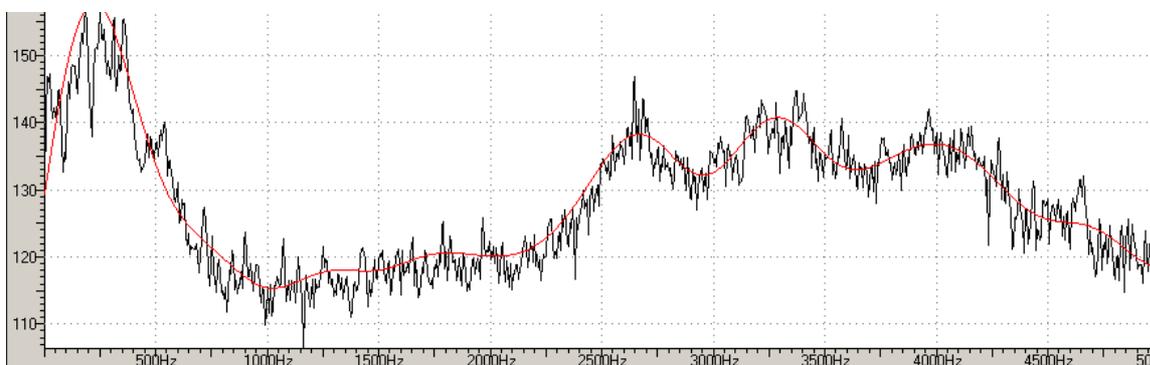
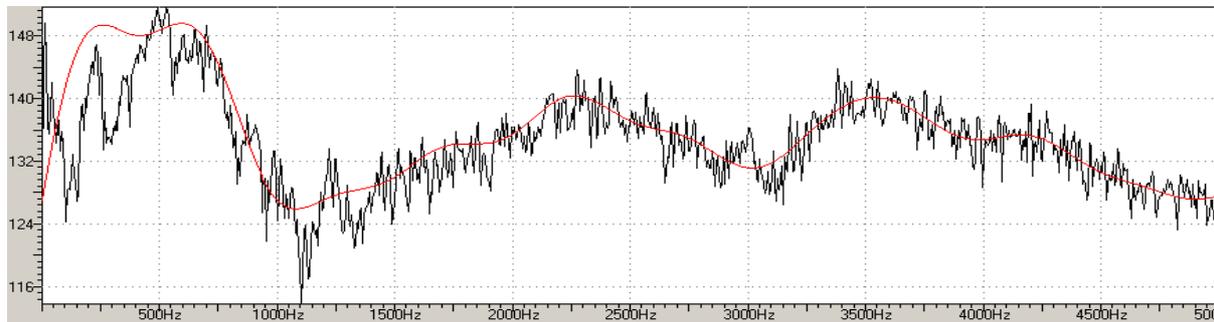


Abb. Sprecher 3 /i/ *Ekel*

Spitzen sind bei 200 Hz, 300 Hz, 500 Hz und eine kleinere Erhöhung noch bei 700 Hz. Wenig Energiedichte findet sich von 1000 bis ca. 2250Hz. Ein weiterer erhöhter Bereich ist von 2600 bis 4200 Hz.

Die **ekelig geäußerten Vokale /e/** von den Sprecherinnen 2 und 4 sind sehr gut erkannt worden:



Sprecherin 2 /e/ Ekel

Die höchste Energiedichte ist bei 500 Hz, eine weitere bei ca. 2300 Hz. Erhöhte Bereiche zeigen sich auch von 2000–3000 Hz und 3400–3900 Hz.



Abb. Spr. 4 /e/ Ekel

Hier erkennt man gut die Spitzenwerte bei 200 Hz und 400 Hz. Dann gibt es noch eine kleinere Spitze bei 600 Hz. Bei 1050 Hz ist die Energie am schwächsten. Ab 1400 Hz steigt die Energie etappenweise wieder an und erreicht einen etwas höheren Bereich zwischen 2300 und 2950 Hz.

Außerdem ist der Vokal /e/ mit der Intention *Ekel* von Sprecher 1 mit 86,4% gut erkannt worden.

Worin gegen das *ekelige /e/* von Sprecher 3 ist **nur von 22,7%** als *Ekel* wahrgenommen worden und kann somit als „Ausreißer“ bezeichnet werden kann. Daher bedarf dieses Beispiel einer näheren Betrachtung:

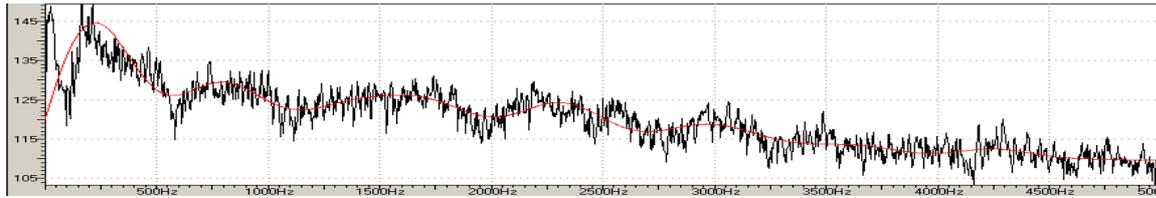


Abb. Sprecher 3 /e/ *Ekel*

Die größte Energiedichte befindet sich bei den unteren und sie schwindet stufenweise Richtung der höheren Frequenzen. Am stärksten ist die Dichte zwischen 250–550Hz mit Spitzen um die 250 Hz.

Anhang 3 Übersicht: Beispiele

Sprecher 1

1_a_angst
8_schm_trauer
10_u_neutral
16_u_ekel
18_i_freude
21_schmu_neutral
22_o_wut
25_i_neutral
26_e_trauer
32_o_trauer
36_a_trauer
38_u_freude
43_a_neutral
48_u_trauer
55_u_wut
62_schmu_angst
67_i_angst
72_e_neutral
73_i_ekel
83_o_ekel
85_schmu_freude
96_a_wut
97_e_angst
108_i_wut
111_schmu_wut
112_o_neutral
119_schmu_ekel
120_a_freude
127_o_angst
128_u_angst
132_e_freude
134_a_ekel
137_e_ekel
138_e_wut
140_o_freude
144_i_trauer

Sprecherin 2

2_schmu_ekel
9_a_angst
14_u_freude
15_u_wut
17_e_trauer
20_i_wut
29_schmu_angst
30_o_neutral
34_e_freude
35_i_ekel
37_schmu_neutral
47_e_wut
49_o_ekel
59_a_wut
61_o_angst
66_i_freude
71_schmu_wut
74_a_neutral
76_i_trauer
77_o_trauer
84_a_freude
87_u_trauer
93_u_angst
94_schmu_freude
95_i_neutral
98_a_ekel
101_u_neutral
102_e_neutral
109_a_trauer
110_o_wut
113_e_angst
117_schmu_trauer
118_u_ekel
121_i_angst
135_o_freude
139_e_ekel

Sprecher 3

3_u_angst
4_schmu_freude
7_a_wut
12_e_freude
19_o_wut
27_u_ekel
33_o_neutral
40_i_wut
41_schmu_neutral
45_schmu_ekel
51_e_neutral
56_e_trauer
60_i_freude
63_o_angst
64_schmu_wut
68_a_freude
70_e_ekel
79_schmu_trauer
80_i_angst
82_a_angst
86_a_neutral
90_u_trauer
103_u_wut
104_o_ekel
105_u_neutral
106_schmu_angst
114_u_freude
115_o_trauer
122_a_trauer
123_e_wut
125_o_freude
126_i_trauer
130_i_ekel
133_a_ekel
141_i_neutral
143_e_angst

Sprecherin 4

5_o_ekel
6_i_neutral
11_o_wut
13_a_neutral
23_u_freude
24_e_freude
28_e_wut
31_schmu_trauer
39_e_trauer
42_u_ekel
44_schmu_wut
46_a_freude
50_o_angst
52_a_ekel
53_i_freude
54_i_angst
57_o_freude
58_schmu_neutral
65_e_neutral
69_u_trauer
75_a_angst
78_i_ekel
81_e_ekel
88_schmu_angst
89_u_neutral
91_o_neutral
92_i_wut
99_schmu_freude
100_a_wut
107_u_wut
116_u_angst
124_e_angst
129_i_trauer
131_a_trauer
136_schmu_ekel
142_o_trauer

Erklärung: /schmu/ ist hier die Abkürzung für das lautpoetische Wort

Anhang 4 Curriculum Vitae

CURRICULUM VITAE

Maria Michaela LEGENSTEIN
Hofkirchen 29
4300 St. Valentin
Tel. 0699 / 17778282
e-mail: mia@sieclelibre.at
web: www.sieclelibre.at (Seite wird gerade umgestellt)
geb. am 9. 10. 1979 in Steyr / Oberösterreich

Ausbildung:

Seit SS 2008	Universität Wien: Studium der Musikwissenschaft voraussichtliches Ende April 2013 (Magistra)
SS 2009	Université Sorbonne Paris: Studienaufenthalt
08/2006 – 02/2007	SAE International Institute Vienna; Zertifikat: Electronic Music Producer
2007 – 2009	Schule für Dichtung Wien: Klassen „Auditive Poesie“ (Lehrer: Gerhard Rühm); „Dakini Poetics“ (Lehrerin: Anne Waldman); „Nichts ist wichtig“ (Lehrerin: Isabeela Beumer); „voice & sound“ (Lehrerin: Sainkho Namtchylak)
09/1999 – 01/2002	Universität Wien; Studium Publizistik mit Fächer- kombination aus Ethnologie, Französisch, Theater- Film- und Medienwissenschaften
06/1999	HLA für Kultur- und Kongressmanagement Steyr; Matura u. Diplom (Kultur- u. Kongressassistentin)

Unterricht (Privatstunden, diverse Institute, Musikschule) in:
Rhetorik, Schauspiel, Stimme & Gesang, Yoga, Tanz (Ballet, Modern- und Jazzdance,
Rock'n'Roll Akrobatik), Performance, Harmonielehre und Klavier, Audio-Programme (z. B.
Ableton Live), Querflöte, Schlagzeug, Basiskenntnisse in diversen Musikinstrumenten (wie
Trommeln, Bassgitarre, etc.)

Sprachen:

Englisch	sehr gut in Wort und Schrift
Französisch	gut in Wort und Schrift
Italienisch, Spanisch	Basiskenntnisse
Latein	kleines Latinum

Besondere Kenntnisse:

Audioprogramme (Logic, Ableton Live, Peak, Traktor, Serato etc.),
Basiskenntnisse in Videoprogramm (Final Cut), Performance-Techniken und das
entsprechende technische Know-How (wie Bedienung von speziellen Mischpulten),
Home-Studio Bereich (Mix-Down, etc.), Recording (Home & Field);
Textverarbeitung, Büromanagement, Archiv (Verwaltung, Systematisierung), Datenbank-
Verwaltung, Erstellen von Texten (Presse, Ankündigungen, Aussendungen), Schriftführung,
Organisation von Pressekonferenzen, Veranstaltungen, Interviews, KünstlerInnenbetreuung,
Konsultation, Recherche, hohe soziale Kompetenz, Teamfähigkeit. Flexibilität etc.

Arbeitserfahrung:

Seit 2002	selbständige internationale Arbeit als DJ, Sound-Poetin, Sound Designer, Künstlerin, (siehe website: www.sieclelibre.at)
seit 2010	Organisation v. Benefizveranstaltungen mit dem Titel: „Models Support“ für Projekte in Afrika
seit 2001	Model für Print/Shows/TV/Film und Kunst (z.B. Vanessa Beecroft) national und international (mehrere Monate Aufenthalte in Athen, Barcelona, Mailand, New York, Madrid, Sydney, Paris)
08/2002 – 08/2012	Schule für Dichtung Wien (www.sfd.at): freelancer; wissenschaftliche Mitarbeit, Archiv, Büro- tätigkeit, Projektassistenz (u. a. Redaktion Publikation: „Falco:Lyrics complete“), Presse- arbeit, Veranstaltungsorganisation
2009/2010	Schule für Dichtung Wien: Lehrerin Klasse „Akustische Poesie“ (www.acoustic-poetry.net) mit Cd-Publikation „Sound Poetry III)
2010	Reportage über Paris inkl. Interviews (Sparten Musik, Kunst, Architektur, Mode, Literatur) f. d. Schweizer Magazin „Adress“
2009	Sound:Frame Festival 2009 Vienna: Leiterin v. Workshop „DJ&Voice“
2006 – 2008	Volksoper Wien: Tänzerin i. d. Oper „Zauberflöte“

09/2003 – 01/2005

Plaza Record Store Vienna: Shop-Assistant

09/2000 – 01/2001

Universität Wien: Tutorin Mediengeschichte

09/1999 – 09/2000

Boomerang Freecards Wien: Bereich Kunstkarten
u. Veranstaltungen (z.B. Organisation Vernissage in
Kunsthhaus Wien)

Hobbies:

Yoga, Komposition (Musik), Kunst (Genuss und
eigene Projekte), Tanz, Natur, wandern, reiten,
schwimmen, singen, lesen, schreiben, Dokumentar-
Filme (Umwelt, Menschenrechte, u. a.) ansehen

Anhang 5 Audiofile mit dem Hörversuch

Eine CD findet sich als Beigabe bei der Print-Version.