universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

„Stimme im Fernsehen – Auditive Emotionswahrnehmung im audiovisuellen Medium“

Verfasserin

Bernadette Geißler

angestrebter akademischer Grad

Magistra der Philosophie (Mag. phil.)

Wien, Juni 2009

Studienkennzahl lt. Studienblatt	A 301 295
Studienrichtung lt. Studienblatt	Publizistik und Kommunikationswissenschaft
Betreuer	Univ.-Prof. Dr. Peter Vitouch

Danksagung

Mein Dank gilt Univ.-Prof. Dr. Peter Vitouch und seinen Assistentinnen, Mag. Stefanie Granzner-Stuhr und Mag. Muna Agha, die mir den Weg klar gemacht und mit Rat zur Seite gestanden haben.

Ich widme diese Arbeit meiner Familie, Patricia, Elisabeth und Josef, die mich immer unterstützt und mein Studium erst möglich gemacht hat. Danke für alles. Dankeschön auch an Felix, der mich immer wieder bestärkt und an mich geglaubt hat. Bedanken möchte ich außerdem bei Max, ohne den meine Vision von Fragebogen wohl nicht Realität geworden wäre. Und bei allen Freundinnen und Freunden, die mich lange Zeit kaum zu Gesicht bekommen, diesen Umstand aber mit stoischer Gelassenheit zur Kenntnis genommen haben.

„Television is an invention that permits you to be entertained in your living room
by people you wouldn't have in your home."
(David Frost)

„They say that ninety percent of TV is junk. But, ninety percent of *everything* is junk."
(Gene Roddenberry)

1. Einleitung

Nicht zufällig stammt das Wort Persönlichkeit vom Lateinischen „personare“, das bedeutet hindurchklingen.[1] Ebenso auffällig: Die Gefühlslage eines Menschen wird mit dem Wort „Stimmung“ bezeichnet, eine Antwort kann nicht nur inhaltlich „stimmig“ sein. Schließlich birgt die Sprechstimme für den Zuhörer eine Vielzahl an Informationen, sie lässt Rückschlüsse auf Alter und Geschlecht des Sprechers sowie seine Stimmungslage und Emotionen zu, auch über soziale Komponenten – etwa die Sympathie oder Rangfolge zwischen Gesprächspartnern – und die Motivation der Person gibt die Stimme Auskunft.[2]

Die Wissenschaft ist auf die Schriftkultur fixiert, da das Medium Schrift das Festhalten von Ergebnissen und Ereignissen sowie Wissenstransfer über lange Zeiträume hinweg erlaubt – und das ohne „störende“ Zusatzinformationen, die in der Stimme (über den Sprecher, seine Sprechweise, Körper- und Persönlichkeitsmerkmale) immer mitschwingen.. Die systematische Erforschung der Stimme auf ihre Bedeutung, den Umgang mit ihr und ihre Rezeption hin ist hingegen ein Stiefkind der Wissenschaft – was lange Zeit auch an den ungenügend ausgereiften technischen Möglichkeiten lag, das Phänomen Stimme valide zu untersuchen. Bis heute ist aufgrund dessen zum Beispiel nicht vollständig geklärt, welche Emotionen welche körperlichen Reaktionen auslösen und wie sich diese im Einzelnen auf die Stimmgebung auswirken.

Zahlreiche Autoren äußern bei Vergleichen zwischen verschiedenen Studien zu Stimmgebung und der Wahrnehmung von Stimme Unzufriedenheit mit unwissenschaftlichen Methoden und einer fehlenden Vergleichbarkeit und Überprüfung der Ergebnisse. Die Stimme ist eine flüchtige Kommunikationsform – sie macht es der Wissenschaft aufgrund ihrer Einzigartigkeit und daraus folgend schwierigen Reproduzierbarkeit nicht leicht, sie zu erforschen.

Wo die Stimme im wissenschaftlichen Diskurs lange Zeit ein Schattendasein geführt hat, hat die Medienentwicklung für die Menschheit das Gegenteil bewirkt: Entstehung und Verbreitung von Radio und Fernsehen haben die Stimme in die Gesellschaft zurückgeholt. Göttert[3] nennt diese Entwicklung „zweite Oralität“.

---

[1] Vgl. Geißner (2002), S. 221
[2] Vgl. Gundermann (1994), S. 36
[3] Vgl. Westphal (2002), S. 42

2. Theorie

Bei Sprache und Sprechen unterscheidet man zwischen der Artikulation – den Bewegungen der peripheren Sprechorgane, um die Lautsprache zu formen – sowie der Phonation, also der Stimmproduktion. Unter Phonation versteht man das mit Hilfe des Stimmapparates hervorgebrachte Schallereignis, das der Stimm- und Sprachproduktion als Grundlage dient. Der Mensch zeichnet sich durch die Leistung der Sprache aus, die Kommunikation mit Hilfe des Sprechens – der äußeren Form der Sprache – ermöglicht. Die Sprache umfasst jedoch auch nonverbale Umsetzungen, die Schriftsprache und das Lesen derselbigen. Voraussetzung für sprachliche Kommunikation über die Stimme ist ein funktionierender „Hör-Sprach-Kreis“ [4] bzw. „Sprech-Hör-Kreislauf“[5], der sowohl das Sprechen als auch das Verstehen beinhaltet.

2.1. Die Anfänge der Stimmforschung

Um 490 bis 430 vor Christus beschäftigte sich der griechische Philosoph Empedokles mit der Stimme, auf seinen Lehren aufbauend wurden Stimmtätigkeiten wie Schreien, freies Sprechen oder Vorlesen in die allgemeine Gesundheitspflege integriert. Hippokrates (460 bis 377 vor Christus) beschäftigte sich mit den körperlichen Instrumenten der Stimmgebung, ebenso wie Aristoteles (384 – 322 vor Christus). Der römische Staatsmann und Rhetoriker Cicero (106 – 43 vor Christus) ließ in seinen Rednerschulen in Bezug auf Empedokles’ Lehre auch über Stimmgebung und Ausdruck diskutieren. Als Begründer der Stimmkunde gilt der römische Arzt Claudius Galen (129 – 199 nach Christus), der anatomische Studien am Sprechapparat – vor allem an Schweinen – durchführte. Während die Erkenntnisse im arabischen Raum weitere Verbreitung fanden, stagnierte die Stimmforschung im europäischen Raum nach dem Märtyrertod des Heiligen Blasius im Jahr 316, der für Stimmheilkunde berühmt war. Erst Leonardo da Vinci (1452 – 1519) holte die Stimmforschung mit seinen anatomischen Studien, Zeichnungen und praktischen Versuchen wieder ans Licht. Sie fand anschließend vor allem an italienischen Universitäten ihre Fortführung. Der Wiener Hofrat Johann Wolfgang von Kempelen (1734 – 1804) konstruierte eine Sprechmaschine, mit der er Vokale und Konsonanten erzeugen und diese zu Silben und Sätzen zusammenfügen konnte. Emanuel di García (1805 – 1906) war ursprünglich Sänger, jedoch überanstrengte er seine Stimme und musste daher bereits in jungen Jahren eine andere Laufbahn einschlagen – er widmete sich fortan der Stimmheilkunde und sorgte mit der Erfindung des Kehlkopfspiegels (Laryngoskop) für einen Meilenstein auf seinem Gebiet. Garcías Erfindung wurde jedoch erst von zwei

[4] Vgl. Friedrich, in: Friedich / Bigenzahn / Zorowka (2000), S. 25
[5] Vgl. Geißner (1991), S. 22

österreichischen Ärzten, Ludwig Türck (1810 – 1868) und Johann Nepomuk Czermak (1828 – 1873) in die Medizin eingeführt – der Kehlkopfspiegel wird bis heute verwendet.[6]

2.2. Physikalische Grundlagen

Stimme und Sprache sind physikalisch gesehen Schallereignisse, also Druckwellen, die von einem schwingenden Körper ausgehen. Sie unterscheiden sich in verschiedenen Dimensionen, die jeder Stimme ihre individuelle Note verleihen:

- Frequenz:
  Die Anzahl der Schwingungen pro Sekunde, sie wird in Hertz (Hz) gemessen. Die Frequenz bestimmt die Tonhöhe – je mehr Schwingungen pro Sekunde, desto höher wird der Ton empfunden. Man spricht hierbei von einer psychoakustischen Größe.[7] Geschieht der Schwingungsvorgang zum Beispiel 440 Mal pro Sekunde, entspricht das einer Tonhöhe von 440 Hertz – dem Kammerton A. Erfolgt die Schwingung doppelt so schnell, ist der Ton eine Oktave höher – doppelt so langsam entsprechend eine Oktave tiefer.[8] Der Mensch nimmt Tonhöhenunterschiede allerdings unterschiedlich stark wahr – so werden Unterschiede zwischen hochfrequenten Tönen als geringer eingestuft als solche zwischen tieffrequenten.[9]
- Amplitude:
  Die Weite der Bewegung eines schwingenden Körpers, die die Stärke der Schallwellen angibt. Dieser Schalldruckpegel wird in Pascal (Pa) oder Mikropascal (µPa) gemessen. Zur Messung wird jedoch meist das logarithmische Maßsystem benützt, Einheit ist hier das Dezibel (dB). Eine Verdoppelung des Schalldrucks entspricht einer Zunahme um 6 dB. Den Grenzwert, ab dem ein Ton hörbar ist, bezeichnet man als Hörschwelle, die bei 0 dB liegt.[10]
- Ton:
  Der „Reinton“ bezeichnet in der Physik eine harmonische Sinusschwingung, die lediglich aus einer einzigen Frequenz besteht. In der Natur kommt der Reinton jedoch praktisch nicht vor.[11]
- Klang:
  Wird in der Alltagssprache dem Ton gleichgesetzt, bezeichnet allerdings richtigerweise

[6] Vgl. Mathelitsch / Friedrich (1995), S. 5-8
[7] Vgl. Zorowka / Höfler, in: Friedrich / Bigenzahn / Zorowka (2000), S. 337
[8] Vgl. Mathelitsch / Friedrich (1995), S. 28 - 30
[9] Vgl. Stassen (1995), S. 8
[10] Vgl. Zorowka / Höfler, in: Friedrich / Bigenzahn / Zorowka (2000), S. 337
[11] Vgl. Friedrich, in: Friedrich / Bigenzahn / Zorowka (2000), S. 44

jedes Schallereignis, das aus einer nicht-einfachen Schwingung entsteht. Der Klang (oder die Teiltöne) besteht aus einer Grundfrequenz, oder Grundton, sowie mehreren Obertönen. Während die Tonhöhe durch den Grundton bestimmt wird, entsteht die Klangfarbe durch die Anzahl und Intensität der Teiltöne.[12] Jede Geräuschquelle (auch die menschliche Stimme) hat eine eigene Klangfarbe, das Timbre – so klingt etwa ein gleich hoher Ton aus einer Flöte völlig anders als aus einer Trompete.

- Eigenfrequenz und Resonanz:
  Jeder schwingungsfähige Körper besitzt eine Eigenfrequenz, die entsteht wenn man den Körper in Schwingung versetzt. Länge, Masse und Spannung des Körpers – etwa der Stimmlippen – bestimmen diese Eigenfrequenz. Stimmt die Frequenz, mit der der Körper zum Schwingen gebracht wird, mit der Eigenfrequenz des Körpers möglichst gut zusammen, entsteht nur geringer Widerstand gegen die Schwingungen – das Resultat ist Resonanz.

2.3. Physiologische Grundlagen

Sowohl Stimm- und Sprachproduktion wie auch –erkennung unterteilen sich beim Menschen in einen peripheren (Organe) sowie einen zentralen (Steuerung im Gehirn) Abschnitt. Nur wenn diese physiologischen Grundlagen erfüllt sind, können stimmliche Äußerungen produziert und wahrgenommen, die darin vermittelten Emotionen empfangen sowie Sprachinhalte verarbeitet und hergestellt werden.

2.3.1. Emotionen

2.3.1.1. Gehirn

Für das Erleben von Emotionen sind laut aktuellem Forschungsstand vor allem drei Teile des Gehirns verantwortlich: die basale Schicht – auch „Reptiliengehirn“ genannt-, das Limbische System und das Großhirn.

In der basalen Schicht, dem Stammhirn und Teilen des Zwischenhirns, sind neben den überlebensnotwendigen körperlichen Funktionen wie der Atmung und dem Herzschlag auch Reflexe und Instinkte verankert. Dieser Teil des Gehirns ist außerdem verantwortlich für den Wachheits- und Erregungsgrad des Menschen, und damit einen wichtigen Teil des emotionalen Erlebens.[13]

Das Limbische System unterhalb der Großhirnrinde ist das Zentrum der menschlichen Grundgefühle, besonders von Angst, Wut, Freude, Trauer und Interesse. Sie werden in dieser

[12] Vgl. Friedrich, in: Friedrich / Bigenzahn / Zorowka (2000), S. 44 f.
[13] Vgl. Hülshoff (1999), S. 33 f.

Phase des emotionalen Erlebens eher als Stimmungen wahrgenommen, können dank einer engen Verbindung zur wichtigsten Hormondrüse, der Hypophyse, jedoch blitzschnell körperlich umgesetzt und der Situation angepasst werden. Stresshormone in als bedrohlich empfundenen Situationen etwa lassen Puls und Blutdruck binnen Sekunden ansteigen, die Pupillen weiten sich u.v.m. – der Körper ist dank der körperlichen Wahrnehmung in Form von Gehörtem, Gesehenem oder Gefühltem und der Umsetzung in das Gefühl Angst im Limbischen System bestmöglich vorbereitet auf eine möglicherweise drohende Gefahr.[14] Besondere Bedeutung bei der Emotionswahrnehmung spielt allem Anschein nach ein Teil des Limbischen Systems, die Amygdala. Besonders in diesem Teil des Gehirns wird offenbar entschieden, ob eine Situation emotionale Bedeutung erhält und so von einer „cold cognition" zu einer „hot cognition" wird. Das bedeutet, dass das Erlebte Relevanz erhält, in eine persönliche Beziehung gesetzt und intensiver erlebt wird.[15] Am empfindlichsten reagiert die Amygdala auf auditive Reize.[16]

In der Großhirnrinde werden Sinneseindrücke schließlich bewusst verarbeitet und in kognitive Prozesse wie Denken, Sprechen oder bewusste Bewegung umgesetzt. Für die Steuerung von Emotionen und das Erleben derselben ist vor allem der Frontallappen des Großhirns, das Stirnhirn, verantwortlich. Hier werden Gefühle klassifiziert und können so auch anderen vermittelt werden, außerdem steht mit diesem Verarbeitungsschritt ein größeres Repertoire an Gefühlen zur Verfügung: Neben den primären Emotionen wie Trauer, Wut und Angst werden im Stirnhirn Liebe, Hoffnung, Eifersucht, Neid und andere Gefühle erfahrbar, was oft auch mit einer Abschwächung des Erlebten einhergeht. So können Emotionen im Gegenteil zu der triebhaften, reflexartigen Steuerung im Reptiliengehirn und dem Limbischen System nun verstanden und kontrolliert werden.[17]

Ob Emotionen in einer einzigen Gehirnhälfte entstehen oder verarbeitet werden, ist wissenschaftlich umstritten. Verschiedenen Untersuchungen zufolge könnte das Emotionszentrums in der rechten Hemisphäre sitzen, da diese mehr Verbindungen zur Amygdala unterhält und Emotionen offenbar überwiegend mit der linken Gesichtshälfte ausgedrückt werden. Ein Ansatzpunkt der Kritik an dieser Theorie ist jedoch, dass bei den zugehörigen Untersuchungen vorwiegend negative Emotionen ausgelöst wurden. Einer

[14] Vgl. Hülshoff (1999), S. 34 ff.
[15] Vgl. Merten (2003), S. 86
[16] Vgl. Ebenda, S. 94
[17] Vgl. Hülshoff (1999), S. 36 f.

weiteren Theorie zufolge werden positive Emotionen vorwiegend in der linken Gehirnhälfte verarbeitet, negative in der rechten. Ähnlich aufgebaut sind Theorien, dass Emotionen, die eine Annäherung und soziales Verhalten nach sich ziehen, links beheimatet sind, Emotionen, denen Vermeidung folgt und die als egoistisch beschrieben werden können, dagegen rechts.[18] Zum aktuellen Stand der Emotionsforschung im Bereich der Stimme siehe Kapitel 3.1.

2.3.1.2. Biochemische Botenstoffe

Das emotionale Erleben wird vom Gehirn nicht nur in Gedanken umgesetzt, auch biochemische Botenstoffe kommen zum Einsatz: Neurotransmitter, Hormone sowie Neuropeptide.

Zu den wichtigsten Neurotransmittern, die das emotionale Erleben mitgestalten, gehören Dopamin, Noradrenalin und Serotonin, deren Ursprungsbahnen in verschiedenen Teilen des Gehirns beheimatet sind. Über Neurotransmitter können Nerven- oder Muskelzellen sowie Drüsen angeregt werden. Hormone arbeiten nach dem divergenten System, das heißt eine Drüse schüttet das Hormon aus, das über den Blutkreislauf an alle Rezeptoren verteilt wird. Das ermöglicht dem Körper zum Beispiel, mit einer Ausschüttung von Adrenalin gleichzeitig die Pupillen zu weiten, Puls und Blutdruck steigen zu lassen, die Herzfrequenz zu erhöhen u.v.m. Zu den Neuropeptiden zählt jene Gruppe biochemischer Botenstoffe, die für das Emotionserleben mitentscheidend ist: die Endorphine. Sie wirken ähnlich wie Opiate, sowohl schmerzstillend als auch euphorisierend, da die Rezeptoren in den Belohnungsstrukturen des Limbischen Systems sitzen.[19]

2.3.2. Voraussetzungen für Stimmbildung und Stimmerkennung

2.3.2.1. Gehirn und Nervensystem

Der erste Teil der Stimmbildung und Stimmerkennung findet in Gehirn und Nervensystem statt:

- Im Gehirn findet der erste Teil des Sprechaktes statt, der auf ein psychisches Erlebnis gründet. Die Signalquelle für Primitivlaute, also der Urform der menschlichen Stimme, ist der Hirnstamm. Das Limbische System verleiht der Stimme ihre Einzigartigkeit, dort werden angeborene Lautelemente in Eigenkompositionen umgewandelt.[20] In den

[18] Vgl. Merten (2003), S. 88 ff.
[19] Vgl. Hülshoff (1999), S. 47 ff.
[20] Vgl. Gundermann (1994), S. 35

motorischen Sprachzentren – der Großhirnrinde und dem Frontallappen – erfährt die Reaktion auf das Erlebnis die Umwandlung in Sprache, Verbalisation genannt. [21]

- Auf diese Vorbereitung folgt die Aktivierung der peripheren Artikulations- und Phonationsorgane über die zentralen motorischen Bahnen (die vordere Zentralwindung, die Pyramidenbahn, das extrapyramidale System, das Kleinhirn sowie das verlängerte Mark).[22]
- Dort aktivieren die Nerven das Ansatzrohr, den Kehlkopf sowie das Zwerchfell und den Atmungsapparat – also die Lunge sowie die zugehörigen Muskelgruppen von Brust und Bauch. Bei diesem Vorgang wird schließlich durch die einsetzende Muskeltätigkeit Luft in charakteristische Schwingungen versetzt – Klänge, Geräusche, Stimme und Sprache entstehen.[23] Die Feinmotorik von Lippen, Gaumen und Kehlkopf werden von der linken Gehirnhälfte aus gesteuert – vermutlich, da sie für die zeitliche Abfolge von Bewegungen verantwortlich ist und daher feine, schnelle Abläufe besser koordinieren kann.[24]

Bei der Sprachrezeption, also dem Sprechverständnis, sind ebenfalls unterschiedliche Teile des Körpers involviert:

- Die Schallwellen treffen im Inneren des Ohrs auf das Corti'sche Organ, das mit den Hörnervenfasern verbunden ist. Es wandelt die akustischen Informationen in neuronale Signale um, indem es das Schallereignis in Frequenzen aufteilt und demnach verschiedenen Nervenfasern zuordnet – dieser Vorgang wird als Tonotopie bezeichnet.[25]
- Über den Hörnerv gelangen die Reize in das Zentralenervensystem und den Hirnstamm, wo sie zuerst auf das primäre Hörzentrum treffen. Dort erfolgt der erste aktive Eindruck, der sich auf das Empfinden beschränkt – hier werden zum Beispiel gefährliche Situationen bereits erkannt und Gegenmaßnahmen ergriffen, noch bevor der Mensch die Gefahr aktiv erfasst bzw. das Gehörte bewusst verarbeitet hat. Die Wahrnehmung in Form der Erkennung und Analyse des Gehörten erfolgt erst im nächsten Schritt, wenn die Eindrücke auf das sensorische Sprachzentrum treffen, welches sich in der Großhirnrinde und der Temporo-Parietalregion befindet.[26]

---

[21] Vgl. Friedrich, in: Friedrich / Bigenzahn / Zorowka (2000), S. 26
[22] Vgl. Ebenda, S. 26
[23] Vgl. Ebenda, S. 26
[24] Vgl. Mathelitsch / Friedrich (1995), S. 46
[25] Vgl. Zorowka / Höfler, in: Friedrich / Bigenzahn / Zorowka (2000), S. 329 ff.
[26] Vgl. Friedrich, in: Friedrich / Bigenzahn / Zorowka (2000), S. 26

2.3.2.2. Atmung und Sprechen

Neben der Atmung nützt der Mensch den Atemapparat zur Stimmerzeugung. Luft wird durch das Ansatzrohr und/oder den Mund über den Kehlkopf durch die Luftröhre (Trachea) in die Lungen eingesogen. Dort gelangt sie in die oberen Atemwege – die Bronchien, an deren Ende sich mit den kleinen Lungenbläschen (Alveolen) die unteren Atemwege befinden. An der Oberfläche sind die Lungen von einer zarten Haut überzogen, dem Lungenfell, an der Innenseite des Brustkorbes befindet sich das Rippenfell. Zwischen diesen beiden Häuten befindet sich eine Flüssigkeitsschicht, die den dort herrschenden Unterdruck unterstützt und dafür sorgt, dass sich die Lungen den Bewegungen des Brustkorbes anschließen – denn die Lungen können keine aktiven Bewegungen ausführen. Diese Brustkorbbewegungen erfolgen durch die Zwischenrippenmuskulatur sowie das Zwerchfell, die damit die eigentlichen Atemmuskeln darstellen. Kontrahieren sie, erweitert sich der Brustkorb und der Mensch atmet ein (Inspiration). An der Atmung sind außerdem die Bauch- und Rückenmuskulatur beteiligt sowie die Muskeln des Schultergürtels und des Halses.[27]

Gesteuert wird die Atmung durch das Atemzentrum im Hirnstamm. Rezeptoren ermitteln dort den Kohlenmonoxidgehalt des Blutes und beschleunigen oder verlangsamen die Atmung dementsprechend. Insgesamt haben die Lungen eines gesunden Erwachsenen eine Kapazität von etwa 5 Litern. Bei ruhiger Atmung wird etwa 0,5 Liter davon ein- und ausgeatmet, dieser Anteil wird als Atemruhevolumen oder Atemzugsvolumen bezeichnet. Dazu kommen das inspiratorische sowie das exspiratorische Reservevolumen, jeweils etwa 1,5 Liter, die maximal ein- oder ausgeatmet werden können. Zu diesen insgesamt 3,5 Litern Vitalkapazität kommen 1,5 Liter Restluft, die sich immer in den Lungen befinden – entweicht diese (zum Beispiel bei einer Brustverletzung), kollabieren die Lungen.[28]

In Ruhe beträgt die Atemfrequenz eines Erwachsenen etwa 10 bis 20 Atemzüge pro Minute. Bei der Phonationsatmung, also der Sprech- und Stimmatmung, wird die Ausatmungsphase gegenüber der Einatmungsphase deutlich verlängert. Bei der Ruheatmung beträgt das Verhältnis zwischen Inspirations- und Exspirationsdauer etwa 1 : 1,2 (bis 1,9), bei der Sprechatmung hingegen 1 : 3 bis 4 (im Extremfall sogar bis zu 8). Um die kürzere Dauer des Einatmens zu kompensieren, atmet der Mensch beim Sprechen tiefer ein, außerdem wird der

[27] Vgl. Friedrich, in: Friedrich / Bigenzahn / Zorowka (2000), S. 29
[28] Vgl. Ebenda, S. 31 f.

Luftstrom beim Ausatmen aktiv abgeschwächt, sodass es sogar möglich ist, beim Sprechen weniger Luft auszuatmen als in der Atemruhelage.[29]

2.3.2.3. Die Stimmproduktion

Stimme entsteht laut der myoelastisch-aerodynamischen Theorie (die nach einigen Unstimmigkeiten heute wieder als gültig erachtet wird) durch das Anblasen der gespannten Stimmlippen mit dem Atemstrom.[30] Die Stimmlippen befinden sich im Kehlkopf, wo sie von der Kehlkopfmuskulatur reguliert werden. Zwischen den Stimmlippen befinden sich die Stimmritze (Glottis), die Supraglottis – der Raum oberhalb der Stimmritze – sowie die Subglottis – der Raum unterhalb der Glottis bis zum Beginn der Luftröhre.
Um Stimme zu produzieren, werden die Stimmlippen von der inneren Kehlkopfmuskulatur von der Atmungs- in die Phonationsstellung gebracht – sie werden verschlossen. Durch die anströmende Luft entsteht der subglottische Druck, der nach Ausströmen der Luft abfällt und die Kehlkopfmuskeln – auch durch die Sogwirkung der schnellen Luftströmung – zum erneuten Verschluss der Glottis bewegt. Durch die so entstehende schnelle, regelmäßige Schwingung der Stimmlippen entsteht eine ebenso regelmäßige, dichtere sowie dünnere Luftsäule – die als Schallwelle wahrgenommen wird (primärer Kehlkopfklang)[31].

Neben den Stimmlippen benötigt der Mensch zur Stimmproduktion das Ansatzrohr, auch Vokaltrakt genannt. Als Ansatzrohr werden alle lufthaltigen Räume oberhalb der Stimmritze bezeichnet, die der Klang- und Lautbildung dienen. Dazu zählen die Supraglottis, Kehl-, Mund- und Nasenrachenraum sowie die Mundhöhle, außerdem die Nase und die Nasennebenhöhlen. Die Supraglottis ist auch bei anderen Säugetieren wie Affen vorhanden und beeinflusst beim Menschen Tragfähigkeit und Klangfülle der Stimme mit. Der Mundrachen ist in seiner Geräumigkeit hingegen lediglich beim Menschen vorhanden, am Übergang zur Mundhöhle befinden sich wichtige Teile des Artikulatonsapparates – Gaumensegel, Zäpfchen und Zungenrücken. Durch den Tiefstand des Kehlkopfes beim Menschen wird die Mundhöhle ebenfalls zur Artikulation genützt, differenzierte Lautsprache entsteht durch den Einsatz von Zunge (unter Zuhilfenahme des Gebisses) und der Lippen, sowie dem Gaumensegel. Die Nasenhöhlen dienen als Resonanzraum, der zu- (Nasallaute

[29] Vgl. Friedrich, in: Friedrich / Bigenzahn / Zorowka (2000), S. 32 f.
[30] Vgl. Habermann (2001), S. 69 ff.
[31] Vgl. Friedrich, in: Friedrich / Bigenzahn / Zorowka (2000), S. 38

sind im Deutschen z.B. /m/, /n/) oder bei Orallauten weggeschaltet werden kann – wobei das Gaumensegel die Mundhöhle abschließt.[32]

2.3.2.4. Das Gehör

Die menschlichen Ohren bestehen aus äußerem, Mittel- und Innenohr. Durch die Ohrmuschel werden Schallwellen in den Gehörgang geleitet, wo sie auf das Trommelfell treffen, das Ende des äußeren Gehörgangs. Das Trommelfell fungiert als Membrane, sie gibt die Schallwellen an die drei kleinen Knochen Hammer, Amboss und Steigbügel weiter. Diese verstärken die erzeugten Kräfte um den Faktor 22, allerdings nur bis 80 dB, lautere Höreindrücke werden schützend in abgeschwächter Form weitergeleitet. Im Innenohr der Schnecke, die an das Mittelohr mit der Paukenhöhle und den Knochen anschließt, liegt das Corti'sche Organ, das von der Basilarmembran und vier Reihen mit etwa 24.000 Haarzellen bedeckt ist.[33] Haut und Härchen könnten darauf hindeuten, dass sich das Gehör aus dem Tastsinn entwickelt hat.[34] Der Steigbügel schließt an Schnecke an und bringt die Lymphflüssigkeit, von der das Corti'sche Organ umgeben ist, zum Schwingen. Die dabei entstehenden Spannungen werden auf ungefähr 18.000 Nervenbahnen ins Gehirn weitergeleitet.[35] Neben der Aufteilung in Frequenzen, der Tonotopie, intensiviert das Corti'sche Organ Schwingungen mit niedriger Amplitude aktiv – kochleäre Verstärkung genannt. Damit wird kontrastreicheres Hören ermöglicht.[36]

Der Mensch nimmt Geräusche auch über das Skelett wahr, dabei fungieren die Schädelknochen als Überträger des Schalls an das Innenohr. Diese Art des Hörens ist nicht sehr wirkungsvoll, spielt aber bei höherfrequenten Schallanteilen eine Rolle.[37] So nehmen wir Infraschall – also Bereiche unterhalb von 20 Hz – über den ganzen Körper wahr, was im Extremfall einzelne Organe in Schwingung versetzen und sogar Krankheitssymptome auslösen kann, weshalb auch Infraschall – obwohl nicht aktiv über das Gehör erfassbar – als Lärm gilt.[38]

---

[32] Vgl. Friedrich, in: Friedrich / Bigenzahn / Zorowka (2000), S. 39 ff.
[33] Vgl. Römer (1994), S. 33 ff.
[34] Vgl. Johnson-Laird (1996), S. 321
[35] Vgl. Römer (1994), S. 33 ff.
[36] Vgl. Zorowka / Höfler, in: Friedrich / Bigenzahn / Zorowka (2000), S. 329 ff.
[37] Vgl. Römer (1994), S. 34
[38] Vgl. Wickel / Hartogh (2006), S. 23

Am empfindlichsten ist das Ohr in den Frequenzen zwischen 1.000 und 4.000 Hz[39], wo auch die Haupttöne der Sprechstimme liegen. Die Grundtöne liegen zwar nur bei etwa 80 bis 250 Hz, die Vokalformanten (siehe 2.3.4.4) sind jedoch bei 250 bis etwa 4.000 Hz beheimatet, die Konsonanten darüber – bis zu 8.000 Hz sind möglich.[40] Der hörbare Bereich liegt etwa zwischen 20 und 20.000 Hz. In Bezug auf die Lautstärke hört der Mensch von 0 dB bis etwa 130 dB, wo die Schmerzgrenze liegt.[41]

Nicht nur fremde Schallwellen werden vom Ohr wahrgenommen, ein wesentlicher Prozess bei der Stimm- und Sprachentwicklung ist die auditive Rückkoppelung, das heißt die Fähigkeit des Menschen, seine eigene Stimme wahrzunehmen. Dieses phonatorische Kontrollsystem entwickelt sich erst nach der Geburt mit dem Erlernen der Sprache. Eine Gehörlosigkeit verunmöglicht aus diesem Grund (ohne fremde Hilfe bzw. technische Hilfsmittel) die normale Sprachentwicklung. Selbst Erwachsenen, deren Gehör plötzlich oder mit zunehmendem Alter verloren geht oder beeinträchtigt ist, kann der Verlust dieses Kontrollsystems Probleme bei Stimmgebung (etwa das Schreien statt Sprechen in Zimmerlautstärke bei betagten Personen) und Artikulation bereiten.[42]

2.3.3. Stimme, Sprache und Gehör im körperlichen Entwicklungsprozess

2.3.3.1. Stimme

Bereits bei der Geburt eines Kindes sind die zur Stimm- und Sprachproduktion sowie -wahrnehmung nötigen Organe ausgebildet, sie unterscheiden sich jedoch in Form und Funktion von denen Erwachsener. Die Organe haben zuerst lebensnotwendige Aufgaben inne, zum Beispiel dass ein Kleinkind gleichzeitig atmen und schlucken kann – deshalb liegt der Kehlkopf eines Neugeborenen viel höher.[43] Die erste Lautäußerung eines Kindes ist gewöhnlich der „Geburtsschrei“, auch reflektorisches Schreien genannt. Durch die Abnabelung wird das Kind erstmals vom Stoffwechsel der Mutter getrennt, der eintretende Sauerstoffbedarf wird durch die erste Einatmung gedeckt. Beim Ausatmen folgt schließlich der Geburtsschrei, die Atemzüge danach entfalten die Lungenbläschen und geben gestaute Blut- und Lymphflüssigkeit ab. Alle neugeborenen Kinder weltweit – körperliche Gesundheit

[39] Vgl. Friedrich, in: Friedrich / Bigenzahn / Zorowka (2000), S. 44
[40] Vgl. Zorowka / Höfler, in: Friedrich / Bigenzahn / Zorowka (2000), S. 336
[41] Vgl. Friedrich, in: Friedrich / Bigenzahn / Zorowka (2000), S. 43 f.
[42] Vgl. Ebenda, S. 54 f.
[43] Vgl. Mathelitsch / Friedrich (1995), S. 51 f.

und eine vollständige Reifung zum Geburtstermin vorausgesetzt – schreien bei ihrer Geburt mit 400 bis 500 Hertz.[44]
Diesen Grundton der Stimme behalten die Kleinkinder im Wesentlichen in den ersten Wochen nach der Geburt bei, Schmerzensschreie liegen bei 1.000 bis 2.500 Hz – also exakt in jenem Bereich, in dem das menschliche Ohr am empfindlichsten ist.[45]
Vor der Pubertät liegt die Spanne des Grundtons zwischen 147 bis 698 Hz.[46] Durch die Geschlechtsreife erfährt die Stimme bei Jungen etwa im Alter vom 12. bis zum 16. Lebensjahr durch das starke Wachstum des Kehlkopfes – bedingt durch die Produktion des männlichen Geschlechtshormons Testosteron – ihre entscheidende Veränderung: sie wird merklich tiefer. Die mittlere Sprechstimmlage von Jungen sinkt um eine Oktave, zuvor kippt die Stimme häufig zwischen der tiefen und hohen Stimmlage – im Volksmund „Stimmbruch" genannt. Auch die Stimme von Mädchen sinkt, allerdings nur um eine Terz oder Quart, bedingt durch eine geringe Längenzunahme der Stimmlippen.[47] Im Erwachsenenalter liegt die mittlere Sprechstimmlage beim Mann schließlich zwischen 100 Hz und 150 Hz, bei der Frau zwischen 200 Hz und 250 Hz.[48] Den größten Stimmumfang – zwei Oktaven bis über drei Oktaven bei ausgebildeten Sängern – hat die Stimme zwischen dem 20. und 50. Lebensjahr. Zu Schwankungen kann es insbesondere bei Frauen während ihres Menstruationszyklus oder einer Schwangerschaft kommen.[49] Die Stimme der Frau sinkt im Alter kontinuierlich, bei 80- bis 90-Jährigen liegt die mittlere Sprechstimmlage etwa bei 195 Hz. Auch die Stimme des Mannes wird tiefer, allerdings nur bis etwa zum fünfzigsten Lebensjahr, woraufhin die Grundfrequenz allmählich steigt.[50] Im Alter von 90 Jahren liegt die mittlere Sprechstimmlage eines Mannes etwa zwischen 131 und 147 Hz.[51]

2.3.3.2. Sprache

Die menschliche Stimme ist untrennbar mit der Sprache verbunden, also der Fähigkeit, Gedanken auszudrücken und im Gegenzug die Äußerungen anderer auch zu verstehen. Sprache ist trotz dieser Eigenschaft, die dem Menschen eigen ist, lediglich eine sekundäre Hirnfunktion, die Sprachdominanz entwickelt sich – im Vergleich zu primären Hirnfunktionen wie Hören, Riechen oder dem Tastsinn – erst im Laufe der kindlichen

[44] Vgl. Mathelitsch / Friedrich (1995), S. 55 f.
[45] Vgl. Gundermann (1994), S. 46
[46] Vgl. Ebenda, S. 46
[47] Vgl. Mathelitsch / Friedrich (1995), S. 61
[48] Vgl. Friedrich, in: Friedrich / Bigenzahn / Zorowka (2000), S. 56
[49] Vgl. Mathelitsch / Friedrich (1995), S. 68
[50] Vgl. Gundermann (1994), S. 46
[51] Vgl. Mathelitsch / Friedrich (1995), S. 62 Abbildung

Entwicklung. Als wichtigste Teile des Gehirns gelten in diesem Zusammenhang die Broca-Region in der 3. Stirnwindung als motorisches Sprachzentrum, wo der Entwurf für die Worte entsteht, sowie die Wernicke-Region in der 1. Schläfenwindung als das sensorische Sprachzentrum, wo Wortklangbilder gespeichert werden.[52]
Die Sprachfunktion ist überwiegend in einer Hirnhälfte angesiedelt, wie bei der Rechts- oder Linkshändigkeit entwickelt sich auch die „Lateralität“ (Seitigkeit) der Sprache im Gehirn im Kindesalter. Etwa ab dem dritten Lebensjahr – sobald auch eine Hand bevorzugt wird (diese Entwicklung steht jedoch in keinem kausalen Zusammenhang mit der Sprachlateralität) – wird das Sprachzentrum in einer der beiden Gehirnhälften angesiedelt, ab der Pubertät ist es dort fixiert. Zuvor ist der Spracherwerb zum Beispiel bei einer Gehirnverletzung der betroffenen Areale noch einmal in der anderen Gehirnhälfte erlernbar, nach der Pubertät jedoch nicht mehr.[53]
In 97 Prozent der Fälle – unabhängig von kulturellen oder sozialen Faktoren – ist die linke Gehirnhälfte die sprachlich dominante – wo auch logisch, analytische und rationale Entscheidungen gefällt werden.[54] In der linken Gehirnhälfte wird nicht nur Sprache produziert, auch sprachbezogene Höreindrücke werden hier verarbeitet, zum Beispiel die Syntax. Sprachmelodie sowie Geräusche und Musik hingegen werden in der rechten Gehirnhälfte prozessiert. Emotional gesehen sind neutrale oder positive Gefühle eher in der linken, negative vorwiegend in der rechten Hemisphäre beheimatet.[55]
Frauen sind Männern in verbalen Fähigkeiten etwas überlegen (Männer liegen dafür im räumlichen, geometrischen Bereich voran), was durch eine geringere Lateralisierung der Sprache erklärbar sein könnte – so wird die Informationsverarbeitung beschleunigt.[56]

2.3.3.3. Gehör

Bereits im Mutterleib hat der Mensch Hörerlebnisse, die ersten Ansätze der Ohren bilden sich eine Woche nach der Befruchtung. Am Ende der 16. Schwangerschaftswoche ist – als erstes Sinnesorgan – die Schnecke und damit die körperliche Voraussetzung für das Gehör voll ausgebildet[57], etwa zwei Wochen später beginnen Föten schließlich zu hören. Schon zehn Wochen später haben Ungeborene ein funktionales, auf andere bezogenes Gehör.[58] Bei Föten

[52] Vgl. Denk / Brunner / Bigenzahn, in: Friedrich / Bigenzahn / Zorowka (2000), S. 225 f.
[53] Vgl. Ebenda, S. 226
[54] Vgl. Ebenda, S. 226 ff.
[55] Vgl. Birbaumer / Schmidt, in: Schmidt / Schaible (Hg.) (2006), S. 450
[56] Vgl. Ebenda, S. 454
[57] Vgl. Wickel, / Hartogh (2006), S. 40
[58] Vgl. Chamberlain (1990), S. 55

sind laut Dr. Tomatis[59] die Teile des Gehörs, die für das Hören von Tönen hoher Frequenz benötigt werden, besser ausgebildet als jene für niedrigere Frequenzen, die sich erst in der Pubertät voll ausbilden – gleichzeitig mit dem „Stimmbruch". Abgesehen von Stimmen nimmt das Ungeborene verschiedenste Geräusche wahr, am lautesten von Magen und Darm der Mutter mit bis zu 85 Dezibel, etwa 55 Dezibel laut sind die Herztöne sowie das Pulsieren des Blutes.

Nach der Geburt stellt sich das Gehör um, da nun Schall über die Luft statt das Fruchtwasser übertragen wird. Die Hörschwelle von Babys liegt daher bei etwa 25 bis 30 dB, sie nimmt stetig ab, bis sie im Alter von etwa zwei Jahren wie bei Erwachsenen bei 0 dB liegt. Die Entwicklung der Hörbahnen im Gehirn ist mit 18 bis 24 Lebensmonaten beinahe abgeschlossen, was bei nicht erkannten Hörfehlern nicht nur Probleme des räumlichen Hörens und Sprechens verursacht, sondern auch negative Auswirkungen auf die allgemeine Reifung des Gehirns sowie der Nervenfasern hat.[60]

Bis zum mittleren und fortgeschrittenen Erwachsenenalter verändert sich die Hörfähigkeit nicht, auch hormonelle Veränderungen haben auf sie – etwa im Gegensatz zur Stimme – keinen Einfluss. Als vor allem kulturelles Problem gilt in den Industrienationen die Schwerhörigkeit im Alter, die hauptsächlich als Hörschaden aufgrund von Umwelteinflüssen einzustufen ist. Eine Verminderung des Hörvermögens kann verschiedene körperliche Ursachen haben, etwa die Verkalkung der Gehörknöchelchen, eine schwindende Elastizität des Trommelfells, eine Abnahme des Hörnervs oder das Absterben von Hörzellen – vornehmlich solcher, die auf hohe Töne reagieren.[61]

2.3.4. Die Besonderheiten der Stimme

Jede Stimme ist einmalig, wie Voice Prints beweisen, die der Sprecheridentifikation dienen. Sie können lediglich imitiert werden, identische Kopien sind jedoch nicht bekannt.[62] Die Stimme erhält ihre Eigenheit, ihre Unverwechselbarkeit in verschiedenen Teilen des „Produktionsprozesses". Im Gehirn verleiht das Limbische System der Stimme ihre Einzigartigkeit, dort werden angeborene Lautelemente in Eigenkompositionen umgewandelt.[63] Die Eigenfrequenz entsteht in der Glottis, wo die von der Lunge ausströmende Luft durch das Öffnen und Schließen der Stimmlippen zur Säule wird, die Klang transportiert – den primären Kehlkopfklang. Unter die Grundfrequenz mischen sich

[59] Vgl. Chamberlain (1990), S. 57
[60] Vgl. Wickel / Hartogh (2006), S. 43
[61] Vgl. Ebenda, S. 46 f.
[62] Vgl. Stelzig, in: Lotzmann (Hg.) (1997), S. 9
[63] Vgl. Gundermann (1994), S. 35

bereits hier verschiedene Obertöne, die die Klangfarbe der Stimme bestimmen – sie ist nicht veränderbar. Ganz im Gegensatz zu Tonhöhe und Lautstärke, die ebenfalls im Kehlkopf reguliert werden. Ist die innere Kehlkopfmuskulatur nur wenig gespannt, entstehen tiefere Töne, hohe Töne erfordern langgezogene, stark gespannte Stimmlippen. Hoch zu sprechen ist daher anstrengender und führt bei längerer Dauer zu Stimmermüdung. Die Lautstärke der Stimme wird vom Druck bestimmt, mit dem die Luft auf die Glottis trifft.[64]

2.3.4.1. Stimmlage und -lautstärke

Während des normalen Sprechens kommt die mittlere Sprechstimmlage, auch Indifferenzlage genannt, zum Einsatz, die sich um etwa eineinhalb Oktaven auf und ab bewegt. Sie liegt beim Mann etwa zwischen 100 Hz und 150 Hz (F – H), bei der Frau eine Oktave höher zwischen 200 Hz und 250 Hz (f – h).[65]

Der „Sollwert“ für andauerndes, natürliches und gesundes Sprechen ist die mittlere Sprechstimmlage, auch Indifferenzlage genannt – in ihr ist die Stimmgebung mit dem geringsten Kraftaufwand möglich.[66]

Als Tonhöhen- oder Stimm-Umfang wird jeder Bereich bezeichnet, der den höchst- bis zum tiefstmöglichen Ton umfasst – sowohl beim Sprechen als auch beim Singen. Je nach stimmlicher Übung und Alter unterschiedlich, liegt der Tonhöhen-Umfang im Regelfall bei 1,5 bis 3 Oktaven. Wie die Tonhöhe schwankt auch die Lautstärke beim Sprechen, Minimum sind 50 dB, Maximum 100 bis 110 dB.[67]

Wie gut sich eine Stimme gegen Störlärm behaupten kann, wird mit „Durchdringungsfähigkeit“ oder „Tragfähigkeit“ bezeichnet. Dieser Faktor bestimmt, wie effizient die Stimmproduktion ist. Die Tragfähigkeit wird von der resonatorischen Verstärkung im Ansatzrohr bestimmt.[68]

2.3.4.2. Stimmgattungen

Sängerinnen und Sänger werden je nach Stimmumfang, Klangfarbe und mittlerer Sprechstimmlage in verschiedene Stimmgattungen eingeteilt. Bei den Männern sind dies Bass, Bariton und Tenor, die Frauen unterscheiden sich in Alt, Mezzosopran und Sopran. Diese Stimmgattungen sind jeweils etwa eine Terz voneinander entfernt.

---

[64] Vgl. Friedrich, in: Friedrich / Bigenzahn / Zorowka (2000), S. 46 f.
[65] Vgl. Ebenda, S. 56
[66] Vgl. Siegmüller / Bartels (2006), S. 362
[67] Vgl. Friedrich, in: Friedrich / Bigenzahn / Zorowka (2000), S. 56
[68] Vgl. Ebenda, S. 59

2.3.4.3. Stimmklang

Die Klangfarbe hängt von der Anzahl und Stärke der Obertöne – „Formanten“ genannt – ab, die jeder Stimme ihren individuellen Klang verleihen.[69] Dieses „Timbre“ befähigt uns, eine bekannte Stimme aus einem Stimmengewirr herauszuhören sowie eine Stimme einer einzigen Person zuzuordnen. Der spezifische Stimmklang ist auch noch erkennbar, wenn ein gesprochener Text rückwärts abgespielt wird. Erhöht man dagegen die Abspielgeschwindigkeit um mehr als zehn Prozent, ist der Text zwar noch zu verstehen, das Timbre jedoch geht verloren.[70] Das Timbre ist wissenschaftlich noch nicht vollständig erforscht, die menschliche Wahrnehmung scheint nicht nur von physikalischen Gesetzmäßigkeiten abzuhängen sondern zum Teil psychologisch geprägt zu sein.[71]

2.3.4.4. Formanten

Als Resonanzkörper dient der Stimme das Ansatzrohr, das Frequenzen verstärkt oder abschwächt. Je nach Artikulationsbewegung (Stellung der Zunge, Lippen) werden die Beschaffenheit des Ansatzrohres und so auch die Resonanzfrequenzen verändert, die von ihm verstärkt oder abgeschwächt werden. So entstehen vokal- und zum Teil auch konsonantenspezifische Teiltonmaxima, die Formanten. Erster und zweiter Formant werden vom Gehör die Vokalerkennung benötigt, anhand der weiteren Formanten wird eine Stimme für das Gehör unverwechselbar.[72]

Formant 1 bestimmt, welcher Stimmgattung ein Mensch angehört. Formant 2, der aus mittleren Frequenzen besteht, lässt die Kraft der Stimme oder den Erregungsgrad erkennen. Die Spitzenwerte des Formanten 3 fügen der Stimme einen weinerlichen oder kindlichen Eindruck hinzu. Formant 4 setzt sich aus verschiedenen Lauten zusammen, die vorne am Mund entstehen – eine Mischung aus Zisch- und Reibungsgeräuschen, die bei normaler Sprechweise schwach ausgeprägt sind. Bei emotional aufgeladenem vokalem Ausdruck wie Kichern oder Flüstern hingegen ist Formant 4 stark zu erkennen.[73]

2.3.5. Verarbeitungsprozesse des Gehörs

Das Gehör verfügt im Vergleich zu den Augen in der Aufnahme von Sprache über einen erheblichen Vorteil: Es kann Informationen schneller analysieren und verarbeiten. Wir lesen

[69] Vgl. Friedrich, in: Friedrich / Bigenzahn / Zorowka (2000), S. 56
[70] Vgl Mathelitsch / Friedrich (1995), S. 42 f.
[71] Vgl. Habermann (2001), S. 87 f.
[72] Vgl. Friedrich, in: Friedrich / Bigenzahn / Zorowka (2000), S. 50
[73] Vgl. Sonnenschein (2001), S. 144

viel langsamer als wir zuhören können, da die Augen mehrere Aufgaben haben, sie müssen sowohl den Raum erfassen als auch Bewegungen ausführen und sich daher an einen Zeitablauf halten. Die Ohren hingegen nützen den gesamten Hörbereich auf einmal und verfolgen somit schneller.[74] Das visuelle System hat dafür über einen anderen Vorzug: Der Seheindruck wird etwa 250 Millisekunden länger wahrgenommen, als er tatsächlich andauert – dieser Effekt wird als „Persistenz des Sehens“ bezeichnet.[75]
Einfache Seheindrücke – die nicht die Komplexität des Lesens und Verstehens erfordern – hingegen nehmen wir schneller wahr als wir die zugehörigen Geräusche aktiv erkennen, zum Beispiel ist der Klang beim Klatschen erst kurz nach dem Sehen der vollendeten Handbewegung zu hören. Die Ursache für dieses scheinbare Paradox ist, dass das Ohr Höreindrücke in Form von kurzen Sequenzen zusammenfasst und diese Zusammenstellung analysiert. Während der Zeit bis zum Eintreffen der nächsten Sequenz überprüft das Gehirn das Gehörte minutiös und sammelt sämtliche verfügbaren Informationen.[76] Je nach Art des akustischen Reizes werden verschiedene Hirnareale aktiv, die nicht nur der Wahrnehmung des Gehörten an sich dienen, sondern auch Verbindungen herstellen zu nicht vordergründig akustischen Bereichen des Gehirns – etwa wenn ein Musikstück Erinnerungen auslöst.[77]

2.3.5.1. Art der Signalverarbeitung

Die Ohren besitzt im Gegensatz zu den Augen keine Schutz- oder Schließvorrichtung wie die Augenlider, wir hören also ständig und außerdem in alle Richtungen – allerdings nicht immer aktiv.
Die von den Ohren aufgenommenen Informationen werden über vor allem kreuzweise verlaufende Hörbahnen ins Gehirn gesendet. Dennoch ist auch bei verschiedenen Arten des Zuhörens – von gemütlich bis hin zu Aufmerksamkeit fordernd – keine allgemeingültige Lateralität festzustellen, erst bei der Aktivierung durch Sprache wird bei etwa 90 Prozent der Menschen die linke Gehirnhälfte aktiver tätig als die rechte. Von allen Höreindrücken werden von den höheren Neuronen in den Hörbahnen nur wenige bis zur Hörrinde weitergeleitet – der sogenannte Nutzschall, der anhand von Schallmustern erkannt wird. Zu den wichtigsten übermittelten Reizen gehört die menschliche Sprache. Hintergrundgeräusche oder andere Stimmen als diejenige, der gerade die Aufmerksamkeit zuteil wird, können hingegen praktisch ausgeblendet werden, sodass nur noch diejenigen Höreindrücke vom Gehirn

[74] Vgl. Chion (1994), S. 10 f.
[75] Vgl. Goldstein / Irtel (Hg.) (2008), S. 123
[76] Vgl. Chion (1994), S. 12 f.
[77] Vgl. Gruhn (2005), S. 13

verarbeitet und somit aktiv wahrgenommen werden, die anhand der Schallmuster als wichtig eingestuft wurden.[78]

Durch die links und rechts angeordneten Ohren ist der Mensch zu räumlichem Hören fähig. Auf horizontaler Ebene, weil die Schallwellen je nach Richtung jeweils ein Ohr mit wenigen Millisekunden Verspätung erreichen. Befindet sich die Schallquelle nicht direkt seitlich eines Ohrs, reicht die Zeitverschiebung der Höreindrücke für das Gehirn aus, um die Informationen beider Hörnervenbahnen zu verknüpfen und zu erkennen, aus welcher Richtung ein Geräusch genau stammt. Tatsächlich kann das Gehirn so einzelne Höreindrücke auf bis zu drei Winkelgrade genau orten. Befindet sich die Schallquelle nicht auf gleicher Höhe mit oder hinter den Ohren, wird der Schall durch die besondere Form der Ohrmuscheln je nach Herkunftsort verändert.[79]

Ähnlich den Regeln der Gestaltpsychologie (siehe 2.8.3) gruppiert das auditorische System die ankommenden Reize, was die Verarbeitung vereinfacht. So werden Schallereignisse nach dem Ort eingeteilt, wo sie wahrgenommen werden – auch wenn sie sich langsam bewegen nimmt das Gehör dennoch wahr, dass es sich um eine einzelne Geräuschquelle handelt (etwa ein vorbeifahrendes Auto). Außerdem werden Geräusche zusammengefasst, die sich in Klangfarbe und Tonhöhe ähneln – wodurch zum Beispiel die Unterscheidung zwischen zwei Menschen möglich wird, die gleichzeitig sprechen.[80]

2.3.5.2. Wahrnehmungsgeschwindigkeit

Ein Geräusch wird ab einem Höreindruck von einer Millisekunde wahrgenommen.[81] Ein Geräusch wird so lange als ein einziges wahrgenommen, wie es in einem zeitlichen Abstand von unter etwa drei Millisekunden gehört wird. Liegt zwischen den Höreindrücken ein Abstand von mehr als drei Millisekunden, werden diese als verschiedene Geräusche erkannt. Zum Vergleich: Damit die Augen zwei Bilder als getrennt wahrnehmen, braucht das visuelle System 20 bis 30 Millisekunden.[82]

Unbewusst nehmen wir über das Gehör vor allem Warn- und Alarmsignale auf, die unwillkürlich eine Reaktion nach sich ziehen – zum Beispiel das Ausweichen bei Ertönen der Sirene eines Krankenwagens. Durch den Hirnstamm ausgelöste Reflexe benötigen nur 5 bis 7 Millisekunden, üblicherweise dauert die Verarbeitung eines Höreindrucks bis zur Reaktion

[78] Vgl. Zenner, in: Schmidt / Schaible (Hg.) (2006), S. 308
[79] Vgl. Römer (1994), S. 40 f.
[80] Vgl. Goldstein / Irtel (Hg.) (2008), S. 302 f.
[81] Vgl. Sonnenschein (2001), S. 90
[82] Vgl. Grau, in: Felsmann (Hg.) (2008), S. 39

0,5 bis 1 Sekunde.[83] Die Tonkennzeit – das heißt jene Zeitspanne, die das Ohr benötigt um die Höhe eines Tons festzustellen – hingegen beträgt lediglich zwischen vier und 32 Millisekunden, sie variiert je nach Frequenz und Art des Schallereignisses.[84] Ab einer Länge von 50 Millisekunden wird die Lautstärke wahrgenommen, für die Klangfarbe eines Geräuschs sowie die Erkennung von Konsonanten in der Sprache ist 0,1 Sekunde nötig.[85] Wie sehr das menschliche Gehör auf die Stimme und Sprache programmiert ist, zeigt, dass der Mensch bis zu 30 Phoneme – also die kleinsten Spracheinheiten wie zum Beispiel /r/ – pro Sekunde unterscheiden und damit aktiv wahrnehmen kann. Hingegen können nur 10 musikalische Noten pro Sekunde verarbeitet werden.[86]

2.4. Kognitive und psychologische Grundlagen

2.4.1. Emotionen

2.4.1.1. Einführung

Hülshoff[87] definiert Emotionen als „körperlich-seelische Reaktionen, durch die ein Umweltereignis aufgenommen, verarbeitet, klassifiziert und interpretiert wird, wobei eine Bewertung stattfindet.“ Emotionen wirken demnach zuerst auf den Körper, indem sich ein Reiz auf das vegetative Nervensystem und die Organsysteme auswirkt. Gleichzeitig reagiert der Mensch mit willkürlicher und unwillkürlicher Motorik, wie einer anderen Stimmlage, einem bestimmten Gesichtsausdruck oder der Veränderung der Körperhaltung.

Gefühle zu empfinden ist für den Menschen nicht nur im Umgang mit anderen Menschen zur Kommunikation unabdingbar, um Beziehungen aufzubauen und Empathie zu empfinden. Emotionen und die einhergehenden körperlichen Veränderungen (Mimik, Ausdruck, Verhalten) definieren auch die Identität des Selbst, die beteiligten chemischen Botenstoffe sorgen außerdem für psychische und physische Gesundheit.[88]

Emotionen können unterschiedlich klassifiziert werden: In Bezug auf die Intensität werden sie ruhiger oder erregter empfunden, während die hedonistische oder Valenz-Dimension

[83] Vgl. Wickel / Hartogh (2006), S. 49
[84] Vgl. Stassen (1995), S. 8
[85] Vgl. Sonnenschein (2001), S. 90
[86] Vgl. Ebenda, S. 134
[87] Vgl. Hülshoff (1999), S. 14
[88] Vgl. Ebenda, S. 20

beschreibt, ob ein Gefühl als unangenehm oder als Lustgewinn eingestuft wird.[89] Von wissenschaftlicher Seite wird diese einfache Einteilung jedoch meist als unzureichend angesehen, da Gefühle wesentlich differenzierter erlebt werden. So werden zum Beispiel sowohl große Angst als auch Wut als sehr unangenehm und mit starker Erregung empfunden, die beiden Gefühle werden jedoch äußerst unterschiedlich erlebt.
Zur stärkeren Differenzierung der Valenzdimension kann zusätzlich zwischen der Kontrolle oder Dominanz einer Emotion unterschieden werden, also einer Tendenz der Zuwendung oder der Abwendung. Kontrolliert das Subjekt seine Umwelt, unterscheidet man zwischen „Weg von mir“ und „Hin zu mir“ – also zum Beispiel Ekel und Wut kontra Freude und Wolllust. Wird das Subjekt hingegen von der Umwelt kontrolliert, kommt es zu „Weg von dir“ oder „Hin zu dir“-Emotionen, zum Beispiel Angst und Traurigkeit im Gegensatz zu Liebe und Sehnsucht.[90]

Welche Emotionen als wichtig erachtet werden und wie bzw. wie stark sie (zum Beispiel öffentlich) dargestellt werden dürfen, ist ebenso bedeutend wie die Wert- und Moralvorstellungen einer Gesellschaft. Bei der Vermittlung dieser soziokulturellen Vorgaben spielen neben der Familie, Freunden und Lehrern auch die Medien eine große Rolle, wie diverse Studien zu unterschiedlichen Emotionen nachweisen konnten.[91] Reaktionen auf emotionale Situationen werden ebenso vermittelt wie gesellschaftliche Normen und Regeln zur Darstellung von Emotionen.

2.4.1.2. Emotion, Gefühl, Affekt, Stimmung und Empathie

In der wissenschaftlichen Auseinandersetzung mit dem Thema Emotion wird meist zwischen Emotion und Gefühl unterschieden – was im Normalsprachgebrauch synonym verwendet wird. Merten[92] erklärt die Differenzierung aus der Einseitigkeit des Begriffs „Gefühl“, der lediglich einen Bestandteil der Emotion – das Fühlen – bezeichnet, andere – wie den emotionalen Ausdruck (also auch die Stimmveränderungen) und daraus folgende Handlungen – jedoch unberücksichtigt lässt.
Stassen[93] sieht Affekte als tiefgründige Prozesse, die für emotionales Erleben ebenso verantwortlich sind wie für grundlegende Zustände, etwa Energie, Interesse oder Lebenswille. Affekte werden demnach über Emotionen nach außen getragen, die als soziale Signale der

[89] Vgl. Hülshoff (1999), S. 15
[90] Vgl. Tischer (1993), S. 27
[91] Vgl. Battacchi / Suslow / Renna (1997), S. 79
[92] Vgl. Merten (2003), S. 10
[93] Vgl. Stassen (1995), S. 3

nonverbalen Kommunikation fungieren. Affekte sind jedoch nicht nur interne Zustände, sie können auch von außen eingeleitet, verstärkt oder abgeschwächt werden. Merten[94] zufolge haftet dem Begriff „Affekt“ eine negative Konnotation an, da er unwillkürlich mit einer unreflektierten Reaktion in Verbindung gesetzt wird, der „Affekthandlung“. Einen positiveren Zugang zum Affekt vermittelt Robinson[95]: Demnach beginnt jede Emotion mit einer affektiven Bewertung der geänderten Situation. Dieses erste Urteil ist ungenau und vor allem auf Geschwindigkeit ausgelegt. Darauf folgt die kognitive Verarbeitung des Geschehenen, die wesentlich mehr Zeit in Anspruch nehmen und verschiedene Emotionen nach sich ziehen, beziehungsweise die bereits im Affekt ausgelösten verstärken oder abschwächen kann.
„Stimmung“ wird allgemein als mittel- und langfristige emotionale Reaktion gesehen, die sich nicht direkt auf einzelne Reize bezieht. Stimmungen sind demnach ungerichtet und bilden den Hintergrund für das emotionale Erleben an sich.[96]
„Empathie“ hat einen indirekteren Zusammenhang mit Emotion, der Terminus bezeichnet die Fähigkeit zum Mitgefühl, dem Nachempfinden-Können von Gefühlen, und stellt damit einen betont sozialen Begriff dar. Empathie kann sich allerdings nicht nur im Umgang mit anderen Personen zeigen – das Einleben in fiktionale Charaktere aus Büchern, Filmen oder dem Fernsehen wird als Phantasieempathie bezeichnet.

2.4.1.3. Definitionen des Begriffs „Emotion“

Es gibt verschiedenste wissenschaftliche Definitionen, die sich je nach Autor, Zugangsweise und theoretischer Grundlage unterscheiden. Beispielhaft hier zwei unterschiedliche Ansätze:

Eine der meistzitierten Definitionen ist jene von Kleinginna und Kleinginna, die jene von beinahe hundert Autoren in eine Arbeitsdefinition umgewandelt:

> „Emotion ist ein komplexes Interaktionsgefüge subjektiver und objektiver Faktoren, das von neuronalen/humoralen Systemen vermittelt wird, die (a) affektive Erfahrungen wie Gefühle der Erregung oder Lust/Unlust, bewirken können; (b) kognitive Prozesse, wie emotional relevante Wahrnehmungseffekte, Bewertungen, Klassifikationsprozesse, hervorrufen können; (c) ausgedehnte physiologische Anpassungen an die erregungsauslösenden Bedingungen in Gang setzen können; (d) zu Verhalten führen können, welches oft expressiv, zielgerichtet und adaptiv ist.“[97]

---

[94] Vgl. Merten (2003), S. 11
[95] Zitiert nach Smith, in: Bartsch / Eder / Fahlenbrach (Hg.) (2007), S. 43
[96] Vgl. Tischer (1993), S. 5
[97] Zitiert nach Merten (2003), S. 355

Ebenfalls weite Verbreitung hat die Definition von Oatley und Jenkins gefunden, die Handlungsbereitschaft in den Mittelpunkt rückt:

> „(a) Eine Emotion wird üblicherweise dadurch verursacht, daß [sic]eine Person – bewußt [sic] oder unbewußt [sic] – ein Ereignis als bedeutsam für ein wichtiges Anliegen (ein Ziel) bewertet; die Emotion wird positiv erlebt, wenn das Anliegen gefördert wird, und negativ, wenn es behindert wird. (b) Der Kern einer Emotion ist die Handlungsbereitschaft (readiness to act) und das Bereitstellen (prompting) von Handlungsplänen; eine Emotion gibt einer oder wenigen Handlungen Vorrang, denen sie Dringlichkeit verleiht. So kann sie andere mentale Prozesse oder Handlungen unterbinden oder mit ihnen konkurrieren. Unterschiedliche Arten von Handlungsbereitschaften bedingen unterschiedliche Beziehungen zu anderen Personen. (c) Eine Emotion wird gewöhnlich als ein typischer mentaler Zustand erlebt, der manchmal von körperlichen Veränderungen, Ausdruckserscheinungen und Handlungen begleitet wird.“ [98]

2.4.1.4. Aspekte der Emotionen

Emotionen bestehen, wie oben beschrieben, nicht nur aus dem „Gefühl“. Sie dienen Scherer[99] zufolge der Reizbewertung, Systemregulation, Handlungsvorbereitung, der Kommunikation von Reaktion und Intention sowie der Reflexion und Kontrolle. Diese Funktionen erfüllen Emotionen dank der Zusammensetzung aus kognitiver, neurophysiologischer, motivationaler sowie Ausdrucks- und Gefühls-Komponente. Einigkeit über die Bedeutung der einzelnen Komponenten herrscht im Vergleich diverser Emotionstheorien lediglich darin, dass der motorische Ausdruck Teil einer jeden Theorie ist.

Auf die Entwicklung der Emotionen haben verschiedenste Lebensbereiche Einfluss. So ist die Sozialisation für die Schablonen verantwortlich, anhand derer Gefühle in einer Kultur eingeteilt werden. Ebenso bestimmt diese, welche Situationen Kinder erleben, welche Bindungen sie eingehen und wie sie ihre Gefühle beschreiben lernen (siehe 2.4.1.8). Besonders wichtig für die emotionale Entwicklung sind frühkindliche Erlebnisse, sowie der Umgang mit ihnen. So lernt der Mensch, Gefühle zu differenzieren und in sein Bewusstsein

[98] Zitiert nach Merten (2003), S. 96
[99] Vgl. Merten (2003), S. 15

zu integrieren. Daraus werden verschiedene Reaktionsschemata abgeleitet, die auch von der Art, wie eine Person Informationen verarbeitet, abhängen.[100]

Emotionen messbar zu machen ist auf verschiedene Weise möglich, sie können verbal erfasst werden (wie im empirischen Teil dieser Arbeit), durch Beobachtung des motorisch-expressiven Systems (Mimik, Stimme) aufgezeichnet oder anhand physiologischer Veränderungen (zum Beispiel mit Hilfe von Messungen der Hautleitfähigkeit oder Scans des Gehirns) gemessen werden.[101]

2.4.1.5. Emotionstheorien

Es gibt zahlreiche Emotionstheorien, die Gefühle nach unterschiedlichen Gesichtspunkten ordnen. Die klassische Einteilung unterscheidet zwischen dimensionalem und kategorialem Ansatz. Emotionen können laut dimensionalem Ansatz auf Basis weniger, grundlegender Dimensionen (etwa der Valenz) eingeteilt werden – diese Einordnung beruht vor allem auf Emotionswörtern (siehe 2.4.1.7), was Kritik von verschiedenen Wissenschaftlern eingebracht hat.[102]

Im Gegensatz zur dimensionalen Einteilung steht jene nach Kategorien. Emotionen werden hier als Teil der Evolution verstanden, die sich aus Instinkten gebildet haben und zur Lösung von Problemen in der Umwelt entwickelt wurden. Begründer dieser evolutionsbiologisch begründeten Theorien ist neben Charles Darwin William McDougall. Er sah Emotionen als Weiterführung verschiedener Instinkte – etwa die Furcht als Erleben des Fluchtinstinkts. Kritik erntete McDougall aufgrund fehlender Beweise durch empirische Untersuchungen sowie der Beliebigkeit der aufgeführten Instinkte.[103]

Weitergeführt wurde die kategoriale Einteilung unter anderem in der Differentiellen Emotionstheorie – die Grundlage stammt von Sylvan S. Tomkins, weiterentwickelt von Carroll E. Izard. Diese geht von elementaren Emotionen aus, die sich im Lauf der Evolution gebildet haben und seither unveränderlicher Teil des Menschen sind. Laut der differentiellen Emotionstheorie existieren acht bis zehn Grundemotionen: Interesse, Freude, Überraschung, Verzweiflung, Ärger, Abscheu, Verachtung, Scham, Schuld und Angst.[104] Auf Grundlage dieser Theorie entwickelte Robert Plutchik seine Emotionstheorie, die wiederum Basis der

[100] Vgl. Ulich, in: Ulich / Mayring (2003), S. 108
[101] Vgl. Tischer (1993), S. 13
[102] Vgl. Ebenda, S. 30
[103] Vgl. Merten (2003), S. 54
[104] Vgl. Plutchik (2003), S. 93 ff.

praktischen Untersuchung dieser Diplomarbeit ist und daher im Folgenden näher beschrieben wird.

2.4.1.5.1. Emotion nach Plutchik

Der praktische Teil dieser Arbeit bezieht sich auf die psychoevolutionäre Emotionstheorie von Robert Plutchik. Plutchik hat dabei drei Modelle entwickelt, die miteinander in Verbindung stehen: Das „Structural Model", das „Sequential Model" und das „Derivatives Model". Plutchik selbst schränkt ein, dass die Kategorien seiner Theorie nur in Bezug auf die Emotionssprache verwendbar sind, was ihm Kritik einbrachte. Für den praktischen Teil dieser Diplomarbeit erweist sich Plutchiks Theorie jedoch gerade wegen der expliziten Anwendung der Emotionssprache als besonders geeignet. Alle Modelle bauen auf der Idee der Primäremotionen auf – Angst, Wut, Traurigkeit, meist auch Freude, Liebe und Überraschung.

- **„Structural Model"**

Plutchik zufolge unterscheiden sich Emotionen sowohl durch die Intensität, mit der sie erlebt werden, als auch dadurch, ob und wie sehr sie einander ähneln. So stehen einander Gefühle wie Wut und Ekel deutlich näher als Wut und Freude. Auf dieser Basis stehen sich in Plutchiks Emotionsmodell, das wie das Farbrad von Isaac Newton aufgebaut ist, die gegensätzlichen Emotionen gegenüber.[105]

Plutchik bestimmt acht Dimensionen der Grundemotionen: Freude, Akzeptanz/Vertrauen, Angst, Überraschung, Traurigkeit, Abscheu, Wut und Erwartung/Neugierde. Jeweils zwei dieser Basisemotionen ergeben in seinem „Structural Model" sogenannte „Primary Dyads" (also primäre Zweiergruppen): Freude und Akzeptanz werden zu Liebe, Akzeptanz und Angst zu Unterwerfung, Angst und Überraschung zu Scheu, Überraschung und Traurigkeit zu Enttäuschung, Traurigkeit und Abscheu zu Reue, Abscheu und Wut zu Verachtung, Wut und Erwartung zu Aggressivität, Erwartung und Freude zu Optimismus.[106]

- **„Sequential Model"**

Beim „Sequential Model" geht Plutchik davon aus, dass Gefühle nicht linear ablaufen und im Laufe der Evolution dazu entwickelt wurden, dass sich der Mensch nur innerhalb bestimmter, überlebensnotwendiger Grenzen bewegt – Homöostase genannt. So waren die Gefühle wie Angst oder Wut und die zugehörigen, binnen Sekunden ablaufenden körperlichen und

[105] Vgl. Plutchik (2003), S. 103
[106] Vgl. Ebenda, S. 104 f.

geistigen Veränderungen jahrtausendelang Garant für das Überleben der Spezies Mensch, etwa beim Angriff rivalisierender Gruppen oder der Jagd nach wilden Tieren. Gefühle sind jedoch keine Einbahnstraße – die Stimmung, in die sie einen Menschen versetzen, kann wiederum dessen Einstellungen und Verhaltensweisen verändern.
Ein stimulierendes Ereignis wird laut „Sequential Model" je nach Erkenntnisstand sowie der generelle Gefühlslage und des Erregungsgrades wahrgenommen und meist von einem Reaktionsimpuls gefolgt, von Muskelspannung über den Gesichtsausdruck bis hin zu Vorbereitungen zur Flucht oder Schreien. Diese Reaktion wiederum hat erneut Einfluss auf die Gefühlslage, den Erregungsgrad usw., allerdings findet die Rückmeldung nun von außen nach innen statt. Diesen Prozess nennt Plutchik „Behavioral homoestatic feedback system", also verhaltensgesteuertes homöostatisches Rückmeldungssystem.[107]

Ein unerwartetes Ereignis wird etwa mit Erstaunen und der Frage „Was ist passiert?" wahrgenommen, ein Gefühl der Überraschung stellt sich ein. Wahrscheinliche Reaktion ist das Stehenbleiben beziehungsweise Unterbrechen der gerade vorgenommenen Tätigkeit. Als Effekt hat der Mensch Zeit, sich zu orientieren, wodurch er sich informiert und seine Sicherheit zurückgewinnt. Der Erhalt eines wertvollen Objekts wird dagegen etwa als Besitz erkannt und vom Gefühl der Freude begleitet. Als Reaktion wird der Mensch das Objekt behalten oder die Tat wiederholen, durch die er das Objekt erlangt hat. Der Effekt: Der Mensch sammelt Ressourcen.[108]

- **„Derivatives Model"**

Plutchiks drittes Modell beschreibt die Zusammenhänge zwischen verschiedenen Konzepten, die voneinander abgeleitet werden können. So ist zum Beispiel das Gefühl Angst gleichbedeutend mit der evolutionären Funktion, die diese Emotion erfüllt: Sie dient dem Schutz. Wird eine Emotion über längere Zeit hinweg in einer Person beobachtet, wird diese weniger als einzelnes Gefühl sondern mehr als Wesensart dieses Menschen begriffen und dementsprechend mit der Sprache der Charakterzüge analysiert: Die Person wird als ängstlich bezeichnet. In der diagnostischen Sprache hingegen gilt dieser Mensch als abhängig oder vermeidend. Um die eigene Person zu schützen, wird das Gefühl der Angst unterdrückt und der Umgang mit dem Gefühl ist die Vermeidung. Diese Derivate stellt Plutchik für alle von ihm bestimmten Basisemotionen dar.[109]

---

[107] Vgl. Plutchik (2003), S. 105 ff.
[108] Vgl. Ebenda, S. 109
[109] Vgl. Ebenda, S. 108 ff.

2.4.1.6. Emotionale Kommunikation

Emotionale Kommunikation setzt sowohl die Wissen um die Vermittlung, Enkodierungskompetenz genannt, als auch das Verstehen, die Dekodierungskompetenz, voraus. Enkodiert wird, indem Hinweisreize ausgesendet werden – etwa durch Veränderung der Grundfrequenz der Stimme. Da der Mensch aber von Natur aus in den meisten Situationen nicht nur ein Signal aussendet, muss der Dekodierer in der Lage sein, die irrelevanten Informationen auszusondern, um Emotionen zu erkennen. Gelingt ihm das, kann er mit dem anderen fühlen – er verfügt über Empathie. Sowohl En- als auch Dekodierung verlaufen meist unterbewusst. Dabei kann es auch zur unwillkürlichen Gefühlsansteckung kommen, indem zum Beispiel die Bewegungen des Anderen imitiert werden. Hier übernimmt der eigentliche Empfänger der emotionalen Kommunikation nach dem Dekodieren das Gefühl des anderen.[110]

In verschiedenen Untersuchungen wurde die spezifische Dekodierungskompetenz beleuchtet – es gilt als erwiesen, dass die Gefühle vertrauter Personen (zum Beispiel Mütter und Kinder), besser erkannt werden als jene unbekannter Menschen.[111]

Von Kompetenz im näheren Umfeld abgesehen, haben verschiedenen Untersuchungen gezeigt, dass Frauen im Allgemeinen besser dazu in der Lage sind, ihre Gefühle auszudrücken, sie verfügen also über eine größere Enkodierungskompetenz. Ein positiver Zusammenhang wurde außerdem mit persönlichen Merkmalen wie dem Selbstwertgefühl, der Extraversion und dem kognitiven Stil festgestellt.[112] Auch in Bezug auf die Dekodierungskompetenz sind weibliche Probanden den Männern oft überlegen – in diversen Studien (vor allem zum Einordnen von Gesichtsausdrücken) – zeigten Frauen ein größeres Verständnis und eine korrektere Einschätzung von Emotionen als Männer.[113] Einen großen Einfluss auf die En- und vor allem Dekodierungskompetenz hat das Geschlecht des Interaktionspartners, wie verschiedene Untersuchungen belegen. Demnach präsentieren sich Männer in Gegenwart von Frauen stärker und weniger emotional, während umgekehrt Frauen ihren Emotionen in Anwesenheit eines Mannes mehr Ausdruck verleihen. Frauen erkennen außerdem gewisse Emotionen, z.B. Ärger, signifikant schlechter, wenn diese von einem Mann ausgedrückt werden.[114]

---

[110] Vgl. Merten (2003), S. 144 ff.
[111] Vgl. Ebenda, S. 148
[112] Vgl. Ebenda, S. 148
[113] Vgl. Ebenda, S. 156
[114] Vgl. Merten (2003), S. 157 ff.

Ein Grund für die veränderte Selbstdarstellung und Enkodierung sind jene Stereotypen, die einzelne Emotionen und Handlungsweisen überwiegend Männern oder Frauen zusprechen. Während Maskulinität im Westen etwa mit Handeln, Aggressivität, Rationalität und Furchtlosigkeit in Verbindung gebracht wird, steht Feminität unter anderem für Emotionalität, Gemeinschaft, Liebe und mitfühlende Wärme. Diese Stereotype spiegeln sich auch in Selbstbeschreibungen des emotionalen Ausdrucks von Männern und Frauen wieder. Interessanterweise nehmen diese angeblichen Unterschiede zwischen den Geschlechtern jedoch rapide ab oder fallen sogar ganz weg, werden indirekte Untersuchungsmethoden angewandt, wie etwa die Beschreibung einer emotionalen Situation – was den Einfluss der Stereotype auf die Wahrnehmung von Emotionen beweist, aber auch, dass die Stereotype keinen oder nur geringen Einfluss auf das tatsächliche emotionale Erleben haben.[115]

2.4.1.7. Emotionswörter

Emotionswörter wirken auf verschiedenen Ebenen, sie spiegeln anhand der Ausdrucksfunktion das psychische Erleben des Sprechers wider, außerdem regulieren sie ebendieses, da Emotionen durch Selbstreflexion steuerbar werden können. Auf der sozialen Ebene erfüllen Emotionswörter die Darstellungsfunktion, da so Situationen für andere nachempfindbar beschrieben werden können. Sie übernehmen außerdem die sozialregulative Appell-Funktion, da sie Reaktionen hervorrufen.[116]

Die dimensionale Einteilung ergibt bei der Untersuchung von Emotionswörtern im Wesentlichen die Kriterien Valenz (Lust versus Unlust, positiv kontra negativ), Aktivität (Erregung oder Ruhe), Potenz (stark oder schwach) sowie Intensität des Erlebens.[117] Kritiker der Untersuchung von Emotionen anhand der zugehörigen Emotionswörter merken an, dass nur Teilaspekte der Emotion beschrieben werden und nicht festgestellt werden kann, ob die Emotion selbst oder das auslösende Ereignis bzw. Objekt von den Untersuchten bezeichnet wird.
Auch die Einteilung von Emotionswörtern an sich bereitet der Wissenschaft Schwierigkeiten. Da Emotionswörter Teil des Sozialisationsprozesses und der interpersonellen Entwicklung sind, sind sie schwierig zu bestimmen. Eine Möglichkeit haben Ortony, Clor & Foss[118] (1987)

[115] Vgl. Merten (2003), S. 151 ff.
[116] Vgl. Tischer (1993), S. 22
[117] Vgl. Merten (2003), S. 20 f.
[118] Vgl. Tischer (1993), S. 9

vorgelegt, sie schlagen das Einsetzen möglicher Emotionswörter in die Sätze „I feel X“ und „I am X“ ein – erst wenn beide Sätze Sinn ergeben, könne ein Wort zur Beschreibung einer Emotion verwendet werden. Kritik an dieser Methode kam hauptsächlich aufgrund der Definition von Emotion als vor allem valenzgeladenem Erleben auf. Schmidt-Atzert[119] überließ die Entscheidung, ob ein Emotionswort auch tatsächlich eine Emotion bezeichnet 22 Psychologie-Studenten und 22 Nicht-Studenten. Sie beurteilten 112 Emotionswörter auf einer Skala von 1 bis 5, ein hoher Wert entsprach dabei dem Empfinden, das Wort gehöre dem Oberbegriff „Emotion“ an. Die höchsten Mittelwerte erreichten dabei unter anderem Freude, Angst, Wut, Verzweiflung, Hass, Eifersucht, Erregung, Ekel, Traurigkeit, Leidenschaft und Liebe.

2.4.1.8. Erlernen des Emotionslexikons

Bezug nehmend auf den weit verbreiteten Ansatz, dass einige Emotionen – etwa Freude, Interesse, Überraschung, Ekel, Wut oder Furcht – angeboren sind, wird angenommen, dass Kinder mit Auftreten dieser Empfindungen auch die entsprechenden Emotionswörter lernen. Kinder werden dabei von ihren erwachsenen Bezugspersonen geleitet, die die Verwendung der Emotionswörter auf ihre Richtigkeit kontrollieren und wenn nötig korrigierend einschreiten. Der Lernprozess verläuft dabei vermutlich von außen nach innen: Das Emotionswort wird in einen Zusammenhang mit der momentanen Situation und dem inneren Empfinden gesetzt. Das Kind lernt mit jedem Mal, in dem es ein Emotionswort hört oder selbst korrekt benützt und von Erwachsenen bestärkt wird, das Emotionswort als äußere Form für seine oder fremde Empfindungen zu verwenden – Battacchi[120] beschreibt diesen Vorgang als „Bezeichnungstransfer“.

Verschiedenen Studien zufolge verstehen bereits zweijährige Kinder den Unterschied zwischen Emotionswörtern wie Fröhlichkeit oder Traurigkeit und können diese den entsprechenden Empfindungen korrekt zuordnen. Mit etwa drei Jahren sind Kinder dazu in der Lage, nicht nur ihre eigenen sondern auch die Emotionen anderer Menschen mit Emotionswörtern auszudrücken. Die korrekte Verwendung von Emotionswörtern ist bis zum Grundschulalter jedoch nicht in allen Fällen gegeben – grundsätzlich werden jene Emotionswörter als erstes erlernt, die oft von der Mutter gebraucht werden. Doch nicht nur die direkten Bezugspersonen haben einen großen Einfluss auf den Erwerb von Emotionswörtern, auch das soziokulturelle Umfeld spielt eine bedeutende Rolle. Die

[119] Vgl. Tischer (1993), S. 53 f.
[120] Vgl. Battacchi (1997), S. 74 ff.

Verwendung der einzelnen Emotionswörter ist von den sozialen und kulturellen Eigenheiten der Muttersprache und des Landes abhängig.[121]

2.4.2. Stimme, Sprache und Gehör im kognitiven Entwicklungsprozess

2.4.2.1. Der Fötus – Stimme im Mutterleib

Die Sprachentwicklung beginnt genau wie die Entwicklung des Gehörs bereits bei Ungeborenen. Ab einem Alter von etwa 25 Wochen reagiert das Ungeborene bewusst auf Hörerlebnisse von außen, es strampelt etwa bei Musik.[122] Die ersten Hörerlebnisse hat der Fötus jedoch schon wesentlich früher, etwa ab der 18. Schwangerschaftswoche. Zahlreiche Studien gehen davon aus, dass Ungeborene im Mutterleib besonders von der mütterlichen Stimme geprägt werden.

So stellte Truby[123] anhand einer Studie mit Aufnahmen der ersten Schreie von Babys fest, dass Föten die Sprache der Mutter lernen und die neuromuskulären Bewegungen der Sprechorgane für die Phase nach der Geburt im Mutterleib üben. Ist die Mutter stumm, taub oder sehr still kann das zur Folge haben, dass das Kind bei der Geburt nicht oder auf merkwürdige Weise weint. Föten hören laut verschiedener Studien nicht nur die Stimme der Mutter besser, sie erinnern sich auch an diese, wie unter anderem ein Versuch von DeCasper[124] im Jahr 1980 zeigte. DeCasper ließ wenige Stunden alte Babys mittels Nuckeln an einem Schnuller entscheiden, ob sie eine auf Tonband aufgezeichnete Geschichte weiter hören wollten. Aus drei verschiedenen Aufnahmen bevorzugten die Babys immer wieder diejenige mit der Stimme der Mutter.

Chamberlain[125] geht aufgrund der Erkenntnisse zum Hörbereich von Föten davon aus, dass das Ungeborene die hohe Stimme der Mutter besser hört als die tiefere des Vaters, außerdem wird die mütterliche Stimme seiner Ansicht nach zusätzlich über die Knochen übertragen und durch die Gebärmutter als Resonanzkörper verstärkt. Wickel und Hartogh[126] widersprechen Chamberlains Ansicht, ihrer Meinung nach wird der Hörbereich durch das Fruchtwasser so weit abgeschwächt, dass Föten nicht zwischen den Stimmen von Frauen und Männern

[121] Vgl. Battacchi (1997), S.76 ff.
[122] Vgl. Chamberlain (1990), S. 58
[123] Vgl. Ebenda, S. 56 f.
[124] Vgl. Zimmer (1998), S. 35
[125] Vgl. Chamberlain (1990), S. 57
[126] Vgl. Wickel / Hartogh (2006), S. 40 f.

unterscheiden können. Dieser Meinung schließt sich auch Westphal[127] an, ihr zufolge ist der Fötus noch nicht dazu in der Lage, zwischen dem Mutterleib und dem umgebenden Raum oder anderen Personen zu unterscheiden: „Diese ersten Erfahrungen der Stimme des Anderen ergeben noch ein diffuses Ganzes“.

2.4.2.2. Schreiphase – der erste Stimmeinsatz

Von der Geburt bis zum Ende des ersten Lebensjahres befindet sich das Kind in der präverbalen, bzw. prälingualen Phase der Sprachentwicklung.[128]
Das Gehör gesunder Neugeborener ist bereits so gut wie das Erwachsener, allerdings nehmen Neugeborene Laute Chamberlain[129] zufolge vermutlich erst ab einer Dauer von mehr als 300 Millisekunden wahr – was viele Umgebungsgeräusche ausschließt. Neugeborene reagieren besonders gut auf Geräusche, die von einer Quelle direkt vor ihnen ausgehen, sie nehmen jedoch bei allen Geräuschen wahr, aus welcher Richtung sie stammen. Schon in den ersten Tagen nach der Geburt fangen sie an, ihren Kopf in jene Richtung zu drehen, aus der ein Geräusch stammt.

Ab dem Alter von zwei Wochen sind Babys bereits dazu in der Lage, zwischen der menschlichen Stimme – die eine beruhigende Wirkung hat – und anderen Geräuschen unterscheiden.[130] Von der Geburt bis etwa zum ersten Lebensjahr können Babys Phoneme, also die kleinsten Spracheinheiten (etwa „ba“ und „pa“) besser unterscheiden als Erwachsene.[131] Bereits unmittelbar nach der Geburt kann ein Kind außerdem verschiedene Sprachen erkennen, es differenziert anhand verschiedenster, auch nicht-muttersprachlicher Lautkontraste.[132]
Zwischen der zweiten und der fünften Lebenswoche beginnt das Baby, basale vokalartige Laute von sich zu geben – vor allem in Ruhephasen mit geschlossenem oder leicht geöffnetem Mund, was die Stimmbänder schont.[133] Ab der 4. bis 5. Lebenswoche kann das Kind Tonhöhe, Lautstärke, Stimmeinsatz sowie Tonhöhen- und Lautstärkenverlauf des Schreiens bewusst verändern, der „modulierte Schrei“ tritt damit erstmals auf.[134]

---

[127] Vgl. Westphal (2002), S. 105
[128] Vgl. Denk / Brunner / Bigenzahn, in: Friedrich / Bigenzahn / Zorowka (2000), S. 244
[129] Vgl. Chamberlain (1990), S. 60
[130] Vgl. Clark / Clark (1977), S. 376
[131] Vgl. Chamberlain (1990), S. 122
[132] Vgl. Siegmüller / Bartels (2006), S. 25 ff.
[133] Vgl. Papoušek / Papoušek, in: Scherer (1982), S. 82
[134] Vgl. Mathelitsch / Friedrich (1995), S. 57

2.4.2.3. Gurren und Lallen

Von etwa der siebten Woche an bis zum 3. Lebensmonat erweitern Gurrlaute das Repertoire des Kleinkindes.[135] Im Alter von etwa zwei Monaten setzt das Kind diese Laute sowie sein Lächeln ein, wenn es eine freundliche Stimme vernimmt – eine wütende hingegen führt laut signifikant öfter zum Weinen. Babys haben also bereits mit zwei Monaten gelernt, die emotionale Klangfarbe der Stimme zu deuten.[136]

Etwa vom 3. bis zum 6. Lebensmonat an wacht ein Kind erstmals bei Geräuschen oder Sprache auf, interessante Laute lassen es aufhorchen.[137] Mit etwa vier Monaten können Babys männliche und weibliche Stimmen unterscheiden.[138] Etwa in diesem Alter reagiert ein Kind außerdem zum ersten Mal auf seinen eigenen Namen, gleichzeitig beginnt es mit marginalem Lallen – das heißt, es probiert alle möglichen Laute aus. Ab dem 6. Monat kann das Kind Lallen reduplizieren, indem es mehrere Silben aus einem Konsonanten und einem offenen Vokal aneinanderreiht (z.B. babababa).[139] Ob dieses Brabbeln als Anfang des Spracherwerbs oder eine vom sinnvollen Sprechen in Sätzen unabhängige Zwischenphase ist, ist wissenschaftlich nicht einwandfrei geklärt, berichten Clark und Clark[140]. Ihrer Ansicht nach könnte das Brabbeln als Übung für Mund und Vokaltrakt sowie die Intonation dienen.

2.4.2.4. Sprachverständnis und erste Worte

Nach der Lallperiode, wo erstmals Stimme und Sprache eingesetzt werden, folgt etwa im Alter von acht Monaten die Phase des Sprachverständnisses, in denen das Kind stiller ist und aktiver auf die Sprache seiner Umgebung hört. In diesem Alter lernt das Kind, auf die Ausdrucksmöglichkeiten der Betonung und Sprachmelodie zu achten und versucht, Laute und Silben aktiv nachzuahmen.[141] Etwa ab Beginn des Sprachverständnisses des Babys fällt es auch den Eltern deutlich leichter, die Schreie ihres Kindes nach Kriterien wie Verlangen, Überraschung oder Hunger einzustufen, wie eine Studie von Ricks[142] aus dem Jahr 1975 ergab.

Zu Beginn dieser Phase steht die Fähigkeit zur Segmentierung von Wörtern mit stark betonter Silbe am Wortanfang – das Kind kann nun erstmals einzelne Wörter aus dem Sprachstrom

---

[135] Vgl. Siegmüller / Bartels (2006), S. 25 ff.
[136] Vgl. Clark / Clark (1977), S. 377
[137] Vgl. Siegmüller / Bartels (2006), S. 25 ff.
[138] Vgl. Clark / Clark (1977), S. 377
[139] Vgl. Siegmüller / Bartels (2006), S. 25 ff.
[140] Vgl. Clark / Clark (1977), S. 389 ff.
[141] Vgl. Denk / Brunner / Bigenzahn, in: Friedrich / Bigenzahn / Zorowka (2000), S. 244 ff.
[142] Vgl. Clark / Clark (1977), S. 389

herausfiltern.[143] Im Alter von einem halben bis zu einem Jahr versteht es die ersten Wörter (z.B. Mama, Papa, Wauwau), es entwickelt Freude an Geräuschen und Musik.[144]
Mit dem 9. Lebensmonat beginnt die Sensitivität für phonotaktische Regelmäßigkeiten der Muttersprache, zu diesem Zeitpunkt ist das Kind dann zu variierendem Lallen fähig. Nun werden verschiedene Silben kombiniert und der gesamte Artikulationsraum genützt. Mit etwa zehn Monaten bis einem Jahr kann das Kind muttersprachliche Lautkontraste deutlich besser unterscheiden als fremdsprachliche. Gleichzeitig passen sich die kindlichen Äußerungen immer mehr dem Lautinventar der Muttersprache an. Etwa einen halben Monat später beginnt das Kind, auch Wörtern mit schwach betonter Silbe am Wortanfang aus dem Sprachstrom herauszufiltern und zu unterscheiden. Mit etwa einem Jahr kann das Kind schließlich seine ersten eigenen Wörter produzieren.[145]
Vom 12. bis zum 18. Lebensmonat zeigt das Kind erstmals bei Aufforderung auf Gegenstände und reagiert auf entfernte Geräusche. Bis zum 24. Lebensmonat lernt es schließlich, einfache Entscheidungsfragen zu verstehen, Bezug auf den eigenen Namen zu nehmen und in der Folge einfache Anweisungen zu befolgen.[146]

2.4.3. Zuhören und Verstehen

Über das Gehör nimmt der Mensch vor allem Informationen über seine Umwelt auf, was willkürlich, vor allem aber unwillkürlich passiert. Das Gehör warnt außerdem bei Gefahr und hilft bei der Orientierung (siehe 2.3.5). Es erfüllt jedoch auch ästhetische Funktionen – der Mensch empfindet Freude bei Musik, die beruhigende Wirkung kann sich auch auf den körperlichen Zustand auswirken. Vor allem aber dient das Gehör dem Menschen zur Kommunikation. da die Ohren quasi ständig „auf Empfang“ sind, trifft nicht nur Watzlawicks berühmtes Zitat, „Der Mensch kann nicht nicht kommunizieren“ zu, der Mensch kann außerdem kaum weghören und dieser Kommunikation damit entkommen – ein Grund, warum schon Flüstern in ruhiger Arbeitsatmosphäre enorm störend wirken kann.[147]

Sonnenschein[148] stellt fest, dass der Mensch über zwei grundlegende Modi beim Hören von Sprache und Stimme verfügt: Es sei demnach fast unmöglich, nicht auf den Sinn des Gesagten zu hören und nur die Stimmeigenschaften zu beurteilen, wenn in der Muttersprache

[143] Vgl. Siegmüller / Bartels (2006), S. 25 ff.
[144] Vgl. Ebenda, S. 149
[145] Vgl. Ebenda, S. 25 ff.
[146] Vgl. Ebenda, S. 149
[147] Vgl. Wickel / Hartogh (2006), S.48 ff.
[148] Vgl. Sonnenschein (2001), S. 137

gesprochen werde. Diese besondere Bedeutung, die das Erfassen von Sinn hat, verliert Sprache jedoch, wenn eine Fremdsprache oder eine computergenerierte Stimme gehört wird. Bei Fremdsprachen versucht der Mensch daher, über das in der Stimme vermittelte Gefühl zur Erfassung des Sinns zu gelangen.

Bei der auditiven Wahrnehmung werden verschiedene Dimensionen unterschieden: Bei der Lokalisation wird der akustische Reiz seiner Schallquelle räumlich zugeordnet. Dazu nötig ist die Fähigkeit zur Figur-Grund-Wahrnehmung, die das Abheben eines Geräuschinhaltes vom akustischen Hintergrund ermöglicht. Dank der Diskrimination unterscheidet das Gehör zwischen spezifischen Lauten, zum Beispiel Wortpaare mit minimaler Abweichung (Haus – Laus). Wortbedeutungen, Synonyme, Doppeldeutigkeiten, Bedeutungsunterschiede, die Deklination von Verben und vieles mehr unterscheidet der Mensch, wenn seine Fähigkeit zum Verständnis zum Tragen kommt. Beim Erkennen und Wiedergeben auditiver Sequenzen, sowie der Einhaltung der richtigen Reihenfolge bei Zahlen und Wörtern kommt die Serialität zum Einsatz. Stellt der Mensch zwischen Gehörtem und Gesehenem oder Tätigkeiten eine Verbindung her – zum Beispiel beim Diktat in der Schule –, nützt er die Fähigkeit zur Integration. Wörter ruft er aus dem Gedächtnis anhand der Wiedergabe ab.[149]

2.4.3.1. Arten des Zuhörens

Chion[150] unterscheidet drei Arten des Zuhörens: Causal, Semantic und Reduced. Am meisten gebraucht der Mensch das „kausale Zuhören“, wodurch Hinweise auf die Ursache oder Quelle für ein Geräusch gefiltert werden. Ist beides klar, können weitere Informationen wahrgenommen und in ein Schema eingefügt werden – etwa die Kategorisierung nach menschlicher, tierischer oder mechanischer Ursache. Meist bleibt es bei dieser allgemeinen Einschätzung. Im Film wird das „kausale Zuhören“ beinahe ständig manipuliert durch die Beziehung zwischen Bild und Ton – dem Zuseher wird die Ursache für ein Geräusch vorgesetzt, ohne dass dies der Wahrheit entsprechen muss. Unterstützt wird diese Verfälschung durch das Phänomen der „Synchresis“ (siehe 2.8.1.2)
Beim „semantischen Zuhören” wird ein Code wie die Sprache benötigt, um Gehörtes zu interpretieren, diese Art des Zuhörens ist also wesentlich herausfordernder und komplizierter. Das „reduzierten Zuhören“ konzentriert sich auf die Charakteristika des Tons, unabhängig von Bedeutung und Ursache. Von Pierre Schaeffer gegründet, sollen dabei einzelne Sounds

[149] Vgl. Diendorfer-Radner, in: Friedrich / Bigenzahn / Zorowka (2000), S. 264 f.
[150] Vgl. Chion (1994), S. 25 ff.

und ihre Charakteristika objektiv beschrieben werden, ohne den Einfluss der persönlichen Wahrnehmung. Im normalen Leben kommt „reduziertes Zuhören“ praktisch nie vor.[151]

Sonnenschein[152] schlägt eine weitere Art des Zuhörens als Unterscheidung vor: „Referential Listening“ meint, dass der Zuhörer sich der Bedeutung eines Geräuschs bewusst ist oder sogar von ihr beeinflusst wird. Das referentielle Hören kann Sonnenschein zufolge sowohl auf Instinkt beruhen, zum Beispiel das Brüllen eines Löwen, als auch kulturell oder von einer bestimmten Epoche geprägt sein, etwa das Klappern von Pferdehufen auf Pflastersteinen. Im Film kann das referentielle Hören durch einen Code bestimmt werden, der einem Geräusch eine Bedeutung gibt.

2.4.3.2. Kommunikation

Geißner[153] weist darauf hin, dass miteinander Sprechen nur insofern möglich ist, als jedem Menschen eine soziale Rolle zukommt – die „Sprechrolle“. Das Denken rund um das Sprechen ist daher an die Hörverstehensfähigkeit geknüpft und orientiert sich an ihr. Eine Theorie der Gesprächsfähigkeit, also des Verstehens des Inhalts der Kommunikation, müsse „folglich die Leistungen erklären, die Sprecher und Hörer mit Hilfe pragmatischer Normen vornehmen, wenn sie wechselseitig verständliche Äußerungen sprechen und verstehen“[154]. Jene nonverbalen Ausdrucksmöglichkeiten beim Sprechen, die Gefühle und Stimmungen kundtun, könnten nicht subjektiv sein, andernfalls „müßten [sic] wir bei jedem Partner neu lernen, welcher – vorwiegend stimmliche – Ausdruck bei ihm Freude oder Ärger, Zuwendung oder Aggressivität bedeutet. Auch bei diesen als Klangfarben bezeichneten Ausdrucksfarben handelt es sich um gelernte, das heißt während der Sozialisation erworbene soziale Ausdrucksmuster.“[155] Zusätzlich zu jenen Sprechmustern, die im Laufe der Kindheit erlernt werden, kommuniziert bzw. dekodiert der Mensch anhand von Hörmustern, die ihm die Interpretation des Gehörten erleichtern und zur Gewohnheit werden.[156]

2.4.4. Rezeption von Stimme

Verschiedene Stimmeigenschaften werden von den meisten Menschen einhellig positiv oder negativ bewertet, mit ihnen werden psychische Eigenschaften des Sprechers assoziiert. Als

[151] Vgl. Chion (1994), S. 28 ff.
[152] Vgl. Sonnenschein (2001), S. 77 f.
[153] Vgl. Geißner (1981), S. 81
[154] Vgl. Ebenda, S. 96
[155] Vgl. Ebenda, S. 120
[156] Vgl. Geißner (1981), S. 123

angenehm wird meist empfunden, wenn die durchschnittliche Stimmhöhe in der Indifferenzlage des Sprechers liegt. Negativ fallen dagegen Sprecher mit zu hoher oder zu tiefer Stimmlage auf, die stark von der Indifferenzlage abweicht. Tiefere Sprechstimmen von Männern werden ebenfalls positiv wahrgenommen, da ihnen Kompetenz und Autorität zugesprochen werden. Kräftige, aber nicht zu laute Stimmen werden für beide Geschlechter als wichtig angesehen, da sie von Dominanz und Vitalität zeugen. Eine dem Gesprächspartner, Raum, Thema oder der Situation nicht angemessene Lautstärke wird dagegen als störend empfunden. Das gleiche gilt für monotone Sprecher mit mangelnder Tonhöhenvariation – die Stimme wird im Extremfall als neurotisch und unsozial bewertet. Unbehaglich fühlen sich Zuhörer bei einer sehr rauen Stimme, da sie mitleiden lässt und mit Wut und Aggression assoziiert wird. Eine leise, flüsternde Stimme hingegen bringen Zuhörer mit schüchternen Menschen in Verbindung, nasale Stimmeigenschaften werden im Deutschen hingegen als arrogant bewertet.[157]

Grundsätzlich gilt: Körperliche Eigenschaften werden oft in der Wahrnehmung mit ähnlichen psychischen Eigenschaften des Sprechers gleichgesetzt. Zu kleine Resonanzräume werden so mit Verschlossenheit und Introvertiertheit assoziiert, klangvolle offene Stimmen hingegen mit Temperament und Aufrichtigkeit.[158]

2.4.4.1. Stimme im Gedächtnis

Auch wenn wir einen Menschen noch nie gesehen haben, also keine Informationen über Aussehen oder sonstige Merkmale besitzen, kann unser Gedächtnis für seine Stimme eine Datei anlegen, um sie wiederzuerkennen. Das passiert auch unbewusst beim „kausalen Zuhören“, wenn wir uns einem Menschen verbunden fühlen, etwa einem Radiomoderator.[159]

Für das Erkennen eines uns bekannten Menschen genügt bereits ein Husten oder sogar eine charakteristische Art und Weise, Luft zu holen, wie Sonnenschein[160] feststellt – unser Gehör ist auf die Individualität der Stimme geprägt. Erkannt wird die Stimme von vertrauten Personen durch sogenannte habituelle Klangfarbenvarianten, die von den körperlichen Merkmalen eines Menschen, seinem Nervensystem, der Körperhaltung, aber auch der Mundart bestimmt werden und das Timbre der Stimme ausmachen.[161]

---

[157] Vgl. Eckert / Laver (1994), S. 162 ff.
[158] Vgl. Ebenda, S. 165
[159] Vgl. Chion (1994), S. 27
[160] Vgl. Sonnenschein (2001), S. 132
[161] Vgl. Habermann (2001), S. 86

Sich eine Stimme in Erinnerung zu rufen, fällt allerdings schwer – Chion[162] geht davon aus, dass der Mensch die Stimme mit dem Sprechen verwechselt: Wir behalten im Normalfall die im Sprechakt vermittelte Bedeutung im Gedächtnis, das Medium Stimme hingegen vergessen wir und erinnern uns erst wieder daran, wenn wir mit der Stimme wieder konfrontiert sind.

2.4.4.2. Vokozentrismus

Chion beschreibt den Menschen und seine Wahrnehmung als „vococentric", die Stimme steht im Mittelpunkt: „Die Anwesenheit einer menschlichen Stimme strukturiert den akustischen Raum, der sie enthält"[163]. Ihm zufolge zählt die Stimme aus diesem Grund nicht einfach zu den Sounds, da sie an erster Stelle der menschlichen Wahrnehmung steht. Ist eine Stimme zu vernehmen, wird das Hören automatisch auf diese Quelle aufmerksam und konzentriert sich darauf, die Bedeutung der Stimme zu erkennen, sie zu orten und wenn möglich zu identifizieren.[164]

2.4.4.3.Stimme sticht hervor

Sonnenschein[165] stellt fest, dass sogar eine einzelne Stimme aus einem Stimmengewirr hervorstechen kann – und zwar dann, wenn sie eine Bedeutung für den Zuhörer hat. Das kann die Stimme eines alten Bekannten in einem belebten Restaurant sein oder auch ein heißer Tipp im chaotischen Umfeld der Börse. Die persönlichen Gewohnheiten beeinflussen ebenfalls, ob eine Stimme aus der Masse heraus sticht, zum Beispiel wenn ein Mensch inmitten fremdsprachiger Personen seine Muttersprache vernimmt.

2.5. Stimme und Sprache

Aus jeder gesprochenen Aussage filtert der Mensch drei Informationen: Neben dem WAS, das sich auf den Inhalt des Gesprochenen und sprecherische Elemente sowie den Akzent und die Satzmelodie (Frage oder Aussage) konzentriert, erkennt der Mensch anhand von Stimme und sprachlichen Eigenheiten, WER spricht und WIE. Das WIE inkludiert neben den Emotionen, die mit einer Aussage vermittelt werden, auch stilistische Besonderheiten wie die Wortwahl oder die Satzlänge.[166]

---

[162] Vgl. Chion (1999), S. 1
[163] Vgl. Ebenda, S.5
[164] Vgl. Ebenda, S.5
[165] Vgl. Sonnenschein (2001), S. 80
[166] Vgl. Geißner (1991), S. 21

2.5.1. Funktionen der Sprache

Sprache erfüllt bei der Kommunikation verschiedene Aufgaben: Mit dem emotiven, bzw. expressiven Aspekt werden Emotionen und Stimmungen transportiert, was über die Stimme passiert. Die konative, bzw. appellative Funktion wendet sich an den Empfänger, sie soll bei ihm eine Veränderung hervorrufen. Über den reverentiellen, denotativen bzw. kognitiven Aspekt werden Informationen vermittelt – was den Inhalt des Sprechaktes meint. Schließlich gibt es auch eine metasprachliche Ebene, die das Zurverfügungstehen eines gemeinsamen Codes benötigt. So können zum Beispiel Freunde wissen, dass mit einer Aussage eigentlich das Gegenteil gemeint ist, da sie sich zuvor – wenn auch nicht immer bewusst oder vorsätzlich – auf diesen Code geeinigt haben. Außenstehende können die metasprachliche Ebene nicht immer korrekt dekodieren. Zur Erkennung der metasprachlichen Ebene kann der emotionale Ausdruck, und damit die Stimme, wesentlich sein.[167]

Die Stimme unterstützt sprachliche Äußerungen durch die Betonung, die die Aufmerksamkeit auf bestimmte Informationen lenkt. Betonung entsteht nicht nur durch die Änderung der Lautstärke, sondern besteht oft darin, dass der Grundton der Stimme höher oder tiefer liegt sowie dass die Dauer von Silben verlängert wird. Auch der Tonfall im Allgemeinen kann wichtige Informationen übermitteln – etwa ob ein Satz zweifelnd, verärgert, ironisch oder gutgelaunt gemeint war.[168]

2.5.2. Reproduktion kontra Flüchtigkeit

Wie Stelzig[169] hervorhebt, hat die Sprache durch die Möglichkeit, sie per Schrift festzuhalten, seit der Grammatik des Sanskrit über die Stimme dominiert. Die Ordnungssysteme der Sprache haben zu Bildung und Logik beigetragen, sodass die kommunikative Aufgabe der Stimme bis heute einen geringeren Stellenwert einnimmt. Zu dieser Entwicklung beigetragen hat, dass Merkmale der Stimme wie Tonhöhe, Klang und Timbre lange Zeit kaum bis nicht messbar waren und zum Teil heute noch schlecht eindeutig bestimmbar sind.

Reproduzierbarkeit und der jederzeit mögliche Zugriff auf Geschriebenes im Vergleich zur Flüchtigkeit der Stimme haben der Sprache zu ihrem unverrückbaren Platz an der Spitze verholfen, unterstützt von der Erfindung des Buchdrucks und der allgemeinen Alphabetisierung. Schrift als geschriebene oder gedruckte Form der Sprache ermöglicht die

---

[167] Vgl. Krah, in: Krah / Titzmann (Hg.) (2006), S. 41 ff.
[168] Vgl. Johnson-Laird (1996), S. 338 f.
[169] Vgl. Stelzig, in: Lotzmann (1997), S. 6

Transformation vom visuellen zum optischen Reiz. Postman[170] schreibt dazu, „dass man die psychologische Wirkung des Überwechselns der Kommunikation vom Ohr um Auge, vom gesprochenen zum gedruckten Wort schwerlich überschätzen kann. Die eigene Sprache in einer derart dauerhaften, reproduzierbaren Form zu sehen erzeugte die denkbar gründlichste Beziehung zu ihr.“ Sich einen gelesenen Text in Erinnerung zu rufen ist ungleich einfacher als sich an eine Stimme und gesprochene Sprache zu erinnern.[171]

2.5.3. Die Rolle der Stimme in der Sprache

Die Stimme hat laut Stelzig[172] im „Signalverbund“ Sprache fundamentale Bedeutung, die jedoch aufgrund ihrer schlechten Messbarkeit durch Flüchtigkeit und Unbewusstheit geschwunden ist. Stimme vermittelt sinnliche Wahrnehmung und kann als „direkteste Ausdruckshandlung“, als unmittelbarste Psychomotorik gesehen werden – bereits bei nicht- und vorsprachliche Signale wie Seufzen, Weinen und Lachen, die genetisch bedingt sind. Die Stimme als „Spiegel der Seele“ vermittelt Persönlichkeitsmerkmale ebenso wie Absichten und Situation der Person, sie erlaubt überdies Identifikation mit dem Sprecher und erleichtert so eine angemessene Antwort. Nicht nur Emotionen, auch die Muttersprache sowie ein Dialekt können Einfluss auf die Stimme haben. Sie verlangen bestimmten Intonationseinsatz, der schwerer zu erlernen oder abzulegen ist als die Sprache an sich.[173]

Während Sprache ein differenziertes System – sozusagen „digitale Sendetechnik“ – ist, das nach abstraktem Denken verlangt, überträgt Stimme durch die Unbewusstheit sozusagen „analog“ und kann daher auch in schwierigen Situationen aufgenommen und verstanden werden. Mehr als die Sprache kann die Stimme Nähe oder Distanz herstellen und so zur Gemeinschaftsbildung oder –auflösung genützt werden, etwa durch Beschwichtigungssignale zur Aggressions- und Angstabmilderung, Hemmsignale zum Verscheuchen oder Rufsignale zur Annäherung an andere sowie Sanftheit zur Vertrauensbildung.[174]

2.6. Stimme und Emotion

Der Mensch ist auf die emotionale Signalvermittlung durch die Stimme angewiesen und auf sie eingestellt – bereits ab einem Alter von zwei Monaten können Babys freundliche von

[170] Vgl. Postman (2006), S. 43
[171] Vgl. Stelzig, in: Lotzmann (1997), S. 7
[172] Vgl. Ebenda, S. 1
[173] Vgl. Ebenda, S. 4
[174] Vgl. Stelzig, in: Lotzmann (1997), S. 10 ff.

wütenden Stimmen unterscheiden und reagieren auf die Situation – Bedrohung bewirkt Schreien, Sicherheit Lächeln und Gurren (siehe 2.4.2.3)

Zwischen Sprecher und Zuhörer entsteht eine emotionelle Bindung durch die situative Hinwendung zur Stimme, da sie eine eigene Aussage zusätzlich zur oder abseits der Sprache vermittelt und auch den Sinn mehrdeutiger sprachlicher Vermittlung klärt. Der „situative Stimmklang“ vermittelt nicht nur das Wesen der Person und ihre Absicht sondern auch ihre Emotionen. Sowohl Aussendung als auch Empfangen von Stimme passiert vor allem unbewusst und intuitiv, was die Stimme weniger manipulierbar macht als die zugehörige sprachliche Mitteilung.[175]

Beim Sprechen werden vor allem über die Stimme – neben anderen Faktoren wie zum Beispiel Wortwahl oder Sprechtempo – emotionale Signale an die Außenwelt abgegeben, was als „kommuniziertes Gefühl“ bezeichnet werden kann. Auf der Zuhörerseite kommt es so zur Einschätzung der Emotionen des Sprechers, außerdem kann ein rein auf die Stimme bezogener „emotionaler Eindruck“ gewonnen werden – zum Beispiel „Er klingt wütend.“ – der allerdings nicht zwangsweise dazu führen muss, die wahren Gefühle des Sprechers zu ergründen.[176]

Scherer[177] kritisierte 1982, dass wenige Untersuchungen von Emotionen in der Stimme höchsten wissenschaftlichen Ansprüchen genügten, außerdem gäbe es im Allgemeinen zu wenig Forschung in diese Richtung. Dank neuer bildgebender Verfahren wie der funktionellen Magnetresonanztherapie hat sich die Forschungslage zwar in den letzten Jahren stark verbessert (siehe aktueller Forschungsstand, Kapitel 3.1)wie Emotionsübermittlung mit der Stimme funktioniert, welche körperlichen Veränderungen dabei eine Rolle spielen und anhand welcher Kriterien Menschen Emotionen aus einer Stimme heraushören, ist allerdings dennoch bis heute unzureichend wissenschaftlich belegt und zum Teil unklar.

2.6.1. Enkodierung von Emotionen durch die Stimme

Johnson-Laird[178] beschreibt den emotionalen Zustand eines Sprechers als eigenen Kommunikationskanal, der alle Aussagen überlagert. Welche stimmlichen Signale dabei

[175] Vgl. Stelzig, in: Lotzmann (1997), S. 10
[176] Vgl. Tischer (1993), S. 8
[177] Vgl. Scherer, in Scherer (1982), S. 297
[178] Vgl. Johnson-Laird (1996), S. 340 f.

genau im Einzelnen Emotionen verraten, ist bis heute wissenschaftlich nicht eindeutig geklärt – es handelt sich um eine Vielzahl von Signalen, die Tonhöhe, Lautstärke, Stimmvolumen, Vokaldauer, Sprechtempo und andere Veränderungen einschließen. Tischer[179] schreibt die bis heute herrschende Unsicherheit bezüglich wissenschaftlich eindeutig gesicherter Kriterien, welche Emotion die Stimme wie verändert, vor allem dem Umstand zu, dass es keine standardisierten Tests zum Vergleich der Stimme gibt – was auch an der Einmaligkeit der Stimme an sich liegt.

Emotionen wirken auf die gegenübergestellten Teile des vegetativen Nervensystems, auf Sympathikus und Parasympathikus, deren Wechselwirkung für die Stabilität aller körperlichen Vorgänge verantwortlich ist. Negativ empfundene Emotionen wie Furcht und Ärger führen zu einer Aktivierung des sympathischen Nervensystems, was zu einer höheren Herzfrequenz und dem Anstieg des Blutdrucks führt. Dies führt laut Williams und Stevens[180] zu einer veränderten Atmung in Frequenz, Volumen und Rhythmus, was Stimme und Sprechweise entsprechend wandelt. Außerdem wirkt sich darauf aus, dass weniger Speichel produziert wird, was den Mund trocken werden lässt. In Entspannung, aber auch bei negativen Emotionen wie Trauer und Resignation wird der Parasympathikus aktiver – dieser Teil des Nervensystems scheint auch zu einem gewissen Grad bewusst steuerbar zu sein. Der Herzschlag verlangsamt sich hierbei und der Blutdruck nimmt ab, was eine höhere Blutzufuhr an die Organe und vermehrte Speichelproduktion zur Folge hat.

Williams und Stevenson[181] haben die Auswirkungen von Sympathikus und Parasympathikus bei emotionalen Inhalten auf die Stimme untersucht, indem sie Schauspieler ein kurzes Theaterstück darstellen ließen: Am schnellsten wurde ihren Ergebnissen zufolge bei neutralen Aussagen gesprochen, gefolgt von Ärger, Furcht und Traurigkeit – wobei die Geschwindigkeit bei Traurigkeit nur halb so groß war wie bei den anderen Emotionen. Williams und Stevenson gehen davon aus, dass die an der Stimmproduktion beteiligten Muskeln bei Traurigkeit durch die Aktivierung des parasympathischen Nervensystems weniger stark angespannt sind, was zu einer flacheren Atmung, dadurch bedingt zu häufigerem Einatmen und mehr Pausen führt. Ärger zeichnet sich laut dieser Untersuchung durch eine höhere Grundfrequenz der Stimme und eine hohe Variationsbreite der Grundfrequenz aus – verursacht wird dies vermutlich durch den intensiveren Einsatz der

[179] Vgl. Tischer (1993), S. 17
[180] Vgl. Williams / Stevens, in: Scherer (1982), S. 308
[181] Vgl. Ebenda, S. 312 ff.

Atem- und Kehlkopfmuskulatur aufgrund der Aktivität des sympathischen Nervensystems. Traurigkeit zeichnet sich durch eine niedrige Grundfrequenz und geringe Veränderungen der Tonhöhe aus. Furcht hingegen ist von starken Veränderungen des Stimmbildes geprägt. Die Forscher schließen daraus auf mangelnde Kontrolle über die an der Stimmbildung beteiligte Muskulatur, ebenso wie der Koordination der Sprechwerkzeuge durch erhöhte Aktivität des sympathischen Nervensystems.
Aus den Studienergebnissen folgern Williams und Stevenson[182] überdies, dass Ärger zu verstärktem subglottalem Druck führt, was höhere Frequenzen (über 1 kHz) in der Stimme nach sich zieht.

Merten[183] und Scherer[184] zufolge ist vor allem eine Änderung der Grundfrequenz durch eine größere Spannung der Stimmlippen in emotional erregenden Situationen nachgewiesen. Außerdem variieren die Formanten, die vom Körper selbst und den Artikulationsbewegungen abhängen (siehe 2.3.4.4), in Tonhöhe und Intensität, was ebenfalls Schlüsse auf die Emotionen zulässt.

Scherer[185] hat 16 Studien – darunter auch jene von Williams und Stevenson – zur stimmlichen Emotionsdarstellung verglichen und trotz großer Unterschiede im Versuchungsdesign zahlreiche Gemeinsamkeiten gefunden. Besonders Ärger und Traurigkeit lassen sich demnach anhand objektiver Kriterien erkennen: Beim Ärger steigt die Grundtonhöhe, der Mensch spricht außerdem lauter und schneller. Im Gegensatz dazu steht die Traurigkeit, wo das Niveau des Grundtons niedrig ist und leiser sowie langsamer gesprochen wird. Beide Dimensionen wurden anhand der üblichen Vorstellung von Ärger und Traurigkeit untersucht, das heißt, dass bei unterdrücktem Ärger und verzweifelter Trauer durch die Änderung des Aktivierungslevels möglicherweise andere Ergebnisse zu erwarten wären, wie Scherer festhält.
Weniger gesichert, da weniger oft untersucht und eindeutig belegt sind andere Emotionen in der Stimme: Freude und Furcht zeichnen sich laut Scherer[186] offenbar ebenso wie Ärger durch einen höheren Grundton aus und das Sprechtempo ist schneller. Die Lautstärke ist bei Freude erhöht, für die Furcht hat sich hier kein eindeutiges Bild aus den Untersuchungen ergeben, ebenso wie für die Gleichgültigkeit, die sich durch schnelles Sprechen aber eine niedrige

[182] Vgl. Williams / Stevens, in: Scherer (1982), S. 320
[183] Vgl. Merten (2003), S. 52
[184] Vgl. Scherer, in Scherer (1982), S. 287 ff.
[185] Vgl. Ebenda, S. 301
[186] Vgl. Scherer, in Scherer (1982), S. 300

Grundtonhöhe auszeichnet. Ebenfalls niedrige Grundtöne sind bei Verachtung und Langeweile zu erkennen, die sich außerdem durch ein langsames Sprechtempo auszeichnen. Bei der Verachtung wird lauter, bei der Langeweile hingegen leiser gesprochen. Durch hohe Variabilität der Grundfrequenz zeichnen sich die Emotionen Freude, Ärger und Furcht aus, Gleichgültigkeit und Traurigkeit sind von geringer Variabilität geprägt.

2.6.2. Vorgetäuschte Emotionen

Lügen oder Täuschungen sind auch anhand der Stimme zu erkennen, vor allem die Stimmlage verändert sich. Diese Diskrepanzen differieren allerdings je nachdem, welchen Zweck die Lüge hat – ob zum Beispiel Wut oder Angst verheimlicht werden sollen. Die Stimmlage verrät einen Täuschungsversuch besser als Gestik und Gesichtsbewegungen wie das Lächeln, jedoch weniger gut als sogenannte „micromomentary Expressions“, also sehr kurze Veränderungen im Gesicht.[187] Welche Veränderungen sich durch Lügen allerdings genau in der Stimme abspielen, ist wissenschaftlich unzureichend belegt. So muss ein Anstieg der Stimmhöhe, der mit einer Täuschung assoziiert werden kann, nicht auf diese abzielen, da er auch auf gesteigerte Erregung im Allgemeinen hindeuten kann. Wer also emotional erregt ist, muss nicht zwangsweise ein Lügner sein – diese Fehldeutung wird als „Othello-Fehler“ bezeichnet.[188]

Schauspieler oder Synchronsprecher, deren Aufgabe es im Grunde ist, Emotionen vorzutäuschen und diese auch in ihre Stimme zu legen, ahmen die Lautmuster verschiedener Emotionen nach. Dieser Prozess findet vor allem in der Großhirnrinde statt, wo Emotionen bewusst verarbeitet und umgesetzt werden.[189]

2.6.3. Dekodierung von Emotionen in der Stimme

Laut Chamberlain[190] erkennen bereits Neugeborene, ob die Aufnahme eines weinenden Babys echt oder computersimuliert ist – die echte bringt sie deutlich mehr aus der Fassung. Verstärkt wird dieser Effekt, wenn die Aufnahme von gleichaltrigen Babys stammt, allerdings ist nicht bewiesen, ob Klangfarbe oder stimmliche Eigenart dazu führen. Diese Forschungsergebnisse könnten jedoch ein Hinweis darauf sein, dass Mitgefühl angeboren ist.

---

[187] Vgl. Merten (2003), S. 139
[188] Vgl. Ekman (1989), S. 70 f.
[189] Vgl. Gundermann (1994), S. 35
[190] Vgl. Chamberlain (1990), S. 60 f.

Das Wissen um die verschiedenen Ausprägungen des Schreiens von Babys scheint ebenfalls intuitiv. Wie Chamberlain[191] berichtet, wurden verschiedenste Schreie – darunter bedingt durch Schmerz, Hunger oder Vergnügen – Hunderten Erwachsenen vorgespielt. Die Untersuchung zeigte, dass fast alle Befragten beider Geschlechter und verschiedensten Alters die Schreie richtig der Ursache zuordnen konnten.

Scherer[192] beschreibt in seiner Zusammenfassung diverser Studien mit Sprachproben Erwachsener, dass negative Emotionen besser erkannt werden als positive. Die Rangfolge verläuft seiner Ansicht nach vom Ärger, der am besten identifiziert wird, über Traurigkeit, Gleichgültigkeit und Freude. Williams und Stevenson[193] kamen in ihrer Untersuchung mit Stimmproben von Schauspielern zu einem anderen Bild – sie haben festgestellt, dass Traurigkeit mit 73 Prozent Genauigkeit am besten erkannt wurde, gefolgt von Ärger mit 51 Prozent und neutralen Aussagen mit 47 Prozent. Furcht mit 27 und Freude mit 28 Prozent Erkennungsgrad lagen am unteren Ende der Skala. Eine mögliche Erklärung für die Ergebnisse könnte in ihrem Untersuchungsdesign liegen – es wurden relativ kurze Sprachproben zur Beurteilung vorgelegt.

Scherer, Koivumaki und Rosenthal[194] haben 1972 in einer Untersuchung nachgewiesen, dass die stimmlichen Veränderungen, die Emotionen nach außen signalisieren, selbst dann noch erkannt werden, wenn die Stimmprobe extrem beschnitten wird und zahlreiche ihrer Bestandteile herausgefiltert werden. Sehr emotionsgeladene Stimmproben wurden Testpersonen in unbearbeiteter Version vorgelegt, einer elektronisch gefilterten sowie einer Probe, die durch „randomized splicing“ zerlegt worden war – die Bestandteile der Sprache wurden also in willkürlicher Reihenfolge neu zusammengesetzt. Außerdem wurde eine Kombination aus letzteren beiden Methoden produziert, diese Sprachproben enthielten am Ende weder Töne aus dem oberen noch dem unteren Frequenzbereich, außerdem fehlten übliche Sprachreize wie Rhythmus und Pausen sowie Intonation. Grundlegende Einschätzungen der Zuhörer in den Dimensionen Valenz, Aktivität und Potenz (siehe 2.4.1.4) waren dennoch durchwegs richtig und stimmten überwiegend mit jenen der unbearbeiteten Sprachproben überein. Die Autoren gehen daher davon aus, dass vor allem die Grundfrequenz

[191] Vgl. Chamberlain (1990), S. 115 f.
[192] Vgl. Scherer, in Scherer (1982), S. 303
[193] Vgl. Williams / Stevens, in Scherer (1982), S. 313 ff.
[194] Vgl. Scherer, in Scherer (1982), S. 304 f.

der Stimme und ihre Variabilität sowie die Lautstärke für das Verständnis von Emotionen in der Stimme von fundamentaler Bedeutung sind.

2.7. Film und Fernsehen

2.7.1. Geschichte des Tonfilms

Ende des Jahres 1895 projizierten die Brüder Lumière ihre ersten Filme für Publikum auf eine Leinwand. Bereits in zwei folgenden Jahren erfuhr der Cinématographe Konkurrenz und dieses sowie ähnliche Vorführgeräte erfuhren eine rasche Verbreitung. In den Jahren bis 1912 verwandelte sich das Kino langsam in einen eigenständigen Wirtschaftszweig.[195] 1927 kam mit „The Jazz Singer" der erste Tonfilm in den USA ins Kino, im Grunde ein Stummfilm mit Ton-Inserts. 1928 kam mit „Lichter von New York" der erste echte Musicalfilm ins Kino. Mit diesen Entwicklungen änderte sich nicht nur die Rezeptionsart, auch die Filmgestalter waren gezwungen, sich auf den Ton einzustellen. So erhielt der Dialog mit einem Schlag besondere Bedeutung, außerdem wurden die Kinomusiker nach und nach entlassen – ihr Schicksal teilten auch zahlreiche Stars der Stummfilmzeit, die den Bedürfnissen der Zuseher nicht mehr genügten. Durch die Einführung eines Standards bei den Lichtton-Systemen fand die technische Umrüstung der Kinos schnell Verbreitung. Der Tonfilm ließ aber nicht nur Stars und Musiker aus Filmen und Kinos verschwinden, er erfand eine neue Sparte im Filmgeschäft: die Synchronisation (siehe 2.9.1).[196]

2.7.2. Anfänge des Fernsehens in Österreich

Am 1. August 1955 wurde in Österreich Fernsehgeschichte geschrieben: die erste Sendung des öffentlich-rechtlichen Rundfunks wurde ausgestrahlt – Fernseher zum Empfang derselbigen waren allerdings noch kaum vorhanden. Vorerst wurden nur an drei Tagen in der Woche Sendungen ausgestrahlt, die am Abend wiederholt wurden. Dabei handelte es sich aufgrund fehlender Finanzen und Technik vor allem um Dokumentar- und Kulturfilme, die vom Staat fürs Kino oder die Fremdenverkehrswerbung produziert worden waren. Außerdem nahmen Studio-Interviews einen Großteil der Sendezeit ein. Empfangbar war Fernsehen in der Anfangszeit nur in Wien, Graz und Linz. Mit Dezember 1955 startete die „Zeit im Bild", die das „Bild des Tages" ablöste – dies wurde aufgrund einer Investition von damals 2.000 US-Dollar pro Monat möglich, womit Weltnachrichten von der Nachrichtenagentur „United Press" (heute: „United Press International") eingekauft wurden. Im Februar 1956 wurde das

[195] Vgl. Monaco (2000), S. 232 ff.
[196] Vgl. Faulstich (2005), S. 55

Programm geändert, es gab nun auch Nachmittagsprogramm, zumeist Sendungen für Kinder und Jugendliche. Das Programm für die Jugendlichen bestand schon in dieser Frühphase des österreichischen Fernsehens aus zahlreichen US-amerikanischen Serien wie „Fury“ und „Lassie“. Mit der Aufnahme des Senderbetriebs auf dem Gaisberg kamen ab August 1956 auch Bürger Westösterreichs in den Genuss des Fernsehens.
Mit 1. Jänner 1957 wurde das „Öffentliche Versuchsprogramm“ durch echtes Fernsehprogramm abgelöst – außer Dienstag wurde nun jeden Tag gesendet. Mit diesem Datum etablierte sich die „Zeit im Bild“ als Start ins Abendprogramm, das nun ein wesentlich differenzierteres Bild bot – Komödien, Quizsendungen, Shows sowie Übertragungen von Opern und Theaterstücken und vieles mehr wurde ausgestrahlt. Waren die Fernsehzuseher mit Jahresende 1955 als 1.420 Personen mit einem „Fernseh-Berechtigungsschein“ noch eine elitäre Gruppe, so waren es Ende 1958 bereits 46.510 Menschen.

Ende 1959 wurde die 100.000er-Grenze überschritten, wodurch sich (auch durch das Einsetzen des Werbefernsehens) die finanzielle Lage des Fernsehens deutlich besserte. 1960 stieg die Zahl der Fernsehbesitzer noch einmal sprunghaft an, sie verdoppelte sich auf 200.000 Personen. Am 11. September 1961 wurde der zweite Kanal des öffentlich-rechtlichen Fernsehens aus der Taufe gehoben – er war anfangs nur an drei Tagen in der Woche im Raum Wien empfangbar. Mit 1. Jänner 1967 trat ein neues Rundfunkgesetz in Kraft, durch die zugehörige Programmreform erfuhr das österreichische Fernsehen auch inhaltliche Änderungen. Damals wurde unter anderem festgelegt, dass der Vorabend der Familie und Unterhaltung gewidmet sein sollte. Am 1. Jänner 1969 wurde im ORF das erste Mal probeweise ein Teil des Ersten Programms in Farbe ausgestrahlt. Ab September 1970 wurde auf beiden ORF-Sendern Vollprogramm gesendet. Mit Ende 1971 war das ORF-Programm durch die Inbetriebnahme eines Senders in Kärnten für alle Österreicher grundsätzlich möglich. 1974 trat ein neues Rundfunkgesetz in Kraft, das den ORF offiziell in eine Anstalt öffentlichen Rechts mit Hörer- und Sehervertretung und unabhängigen Programmintendanten verwandelte.[197]

2.7.3. Fernsehen in Österreich heute

Im Juli 2001 wurde in Österreich das Privatfernsehgesetz verabschiedet, das privates Fernsehen terrestrisch (also via Antenne) empfangbar machte. Seit Juni 2003 ist „ATV“ als einziger privater Fernsehsender in ganz Österreich terrestrisch zu empfangen. Daneben gibt es

197 Vgl. Rest / Amanshauser, in: Fabris / Luger (Hg) (1988), S. 265 ff.

mehrere private TV-Sender, die allerdings lediglich lokal zu empfangen sind – zum Beispiel „Puls TV“, das seit Juli 2004 on Air ist.[198] Mitte Juli 2004 startete mit dem DVB-T-Testbetrieb in Graz eine grundlegende Veränderung des Fernsehens in Österreich, da gemäß einer Verordnung der Europäischen Union die Ausstrahlung analoger TV-Programme spätestens 2015 zu einem Ende kommen soll. DVB-T steht für „Digital Video Broadcasting – Terrestrial“, also digitales terrestrisches Fernsehen. Es ermöglicht vor allem eine effizientere Aufteilung der zur Verfügung stehenden Frequenzen – vier bis sechs Programme statt bisher nur einem – sowie bessere Bild- und Tonqualität. 2010 soll die Umstellung in ganz Österreich erfolgt sein.[199]

Die Programmstruktur des Fernsehens in Österreich lässt sich anhand einer vergleichenden Studie aus dem Frühjahr 2007 vergleichen, wo auf ORF 1 93,4 Prozent der Zeit Sendungen ausgestrahlt wurden, auf ORF 2 81,4 Prozent und auf ATV 62,6 Prozent. Der Rest entfällt auf Werbung, im Falle von ATV in besonderem Maß Teleshopping, und Programmtrailer. Fiktionale Unterhaltung macht mit 59,8 Prozent der Sendezeit auf ORF 1 sowie 34,6 Prozent auf ATV den größten Teil aus, auf ORF 2 nimmt diesen Platz mit 41,4 Prozent der Zeit die Fernsehpublizistik (dazu zählen z.B. Nachrichtensendungen, Magazine, Reportagen sowie Talkformate und Dokumentationen) ein.[200]

Laut Media-Analyse[201] (eine repräsentative Umfrage, durchgeführt wurden 16.029 Interviews mit Personen über 14 Jahren in Österreich) erreichte das Fernsehen allgemein im Jahr 2008 in Österreich eine Reichweite von 63,4 Prozent – zum Vergleich: Radio lag bei 82,1 Prozent, Tageszeitung(en) bei 72,9 Prozent. Der ORF erzielte eine Reichweite von 52,2 Prozent, ATV von 13,5 Prozent. Ausländische TV-Sender erreichten 49,8 Prozent Reichweite. Bei der Nutzungsdauer lag der ORF ebenfalls deutlich voran, wie aus dem AGTT/GfK Teletest[202] hervorgeht: Die TV-Nutzungszeit im Jahr 2008 lag pro Tag beim ORF bei 66 Minuten, bei ATV bei 5, bei Puls4 bei 2, alle anderen Sender kamen auf insgesamt 84 Minuten Nutzungszeit pro Tag. Insgesamt hat im Jahr 2008 also jeder Österreicher im Durchschnitt 156 Minuten pro Tag ferngesehen.

---

[198] Vgl. Medien in Österreich (2006), S. 37 ff.
[199] Vgl. Ebenda, S. 27 f.
[200] Vgl. Steininger / Woelke (Hg.) (2008): S. 32
[201] Vgl. Media-Analyse 2008, URL: http://www.media-analyse.at/
[202] Vgl. Medienforschung ORF, URL: http://mediaresearch.orf.at/index2.htm?fernsehen/fernsehen_nutzungsverhalten.htm

2.7.4. Fernsehserien

Die typische Fernsehserie zeichnet sich durch ihren fiktionalen Charakter aus. Faulstich[203] zufolge machen Serien den größten Teil des Fernsehprogramms aus, das Genre müsse daher „in jedem Fall schon quantitativ als die wichtigste Form aller fiktionalen Fernsehsendungen betrachtet werden". Zu den ersten aus den USA importierten Serien für Erwachsene, die in Österreich ausgestrahlt wurden, zählen „Bonanza", „Raumschiff Enterprise" und „Mit Schirm, Charme und Melone".

Serien können in Episodenserien (mit abgeschlossener Handlung), Mehrteiler (mit fortlaufender Handlung, die an aufeinander folgenden Tagen ausgestrahlt werden), eine Sendereihe (z.B. „Tatort" mit unterschiedlichen Darstellern), Endlos-Serien (à la „CSI"; mit fortlaufendem Handlungsstrang und ohne absehbares, geplantes Ende) sowie „Soaps", also Seifenopern (Vorabendserien, die täglich gesendet werden und meist für Jugendliche produziert werden), eingeteilt werden.[204] Bei allen Serien außer jenen in Episoden liegt das Hauptaugenmerk auf der Verzahnung aufeinanderfolgender Teile, dazu wird besonders bei Seifenopern und Mehrteilern gern ein „Cliff Hanger" eingesetzt – also eine dramatische Entwicklung, die am Ende einer Folge passiert und so zum Wiedereinschalten motivieren soll. Ebenfalls mit Ausnahme jener in Episoden zeichnet Serien aus, dass sie über verschiedene Handlungsstränge verfügen, die mehrere Folgen lang fortgeführt werden.

Eine Fernsehserie grenzt sich durch zahlreiche Eigenheiten von weiteren Serien sowie anderen Produktionsarten ab. Zu den Merkmalen zählen unter anderem eine ähnliche Geschichte und Struktur, verbundene Handlungsstränge und diversifizierte Charaktere. Auch der Hintergrund und die Ausstattung sind charakteristisch, ebenso wie die Erzählweise (etwa mit einem Sprecher aus dem Off).[205]

2.7.4.1.Krimiserien

Das Krimigenre im heutigen Sinn formte sich im 19. Jahrhundert aus, begründet vor allem durch ein unabhängiges Justizsystem und die Vorstellung, dass Menschen Spuren hinterlassen und diese von einer besonders intelligenten Person gelesen werden können. Populär wurde das Genre dank der Massenpressung und dem Zeitschriftenwesen. Auch im Kino setzten sich Kriminalgeschichten bereits in den 1910er Jahren durch. In den 1920ern entstanden die ersten Subgenres, etwa jenes des Gangsterfilms aus den USA. Seitdem haben sich unzählige

[203] Vgl. Faulstich (2008), S 106
[204] Vgl Ebenda, S. 108 f.
[205] Vgl. Burton (2000), S. 116 f.

Unterkategorien herausgebildet, etwa Detektiv-, Polizei-, Gerichts- oder Spionagefilm sowie Thriller.[206]

Ein wahrer Boom an fiktionalen Krimiserien im Fernsehen setzte im deutschsprachigen Raum in den 1970er Jahren mit „Derrik", „Der Kommissar", „Tatort" und anderen ein.[207]

Krimiserien sind sowohl in Episoden als auch als Mehrteiler, Sendereihe und Endlos-Serie zu sehen.

Burton[208] zufolge bestimmt weniger die echte Polizeiarbeit den Inhalt von Polizeiserien – das Bild von den Sicherheitskräften wird vielmehr durch Krimiserien sowie weitere Medien geprägt, auch wenn diese keine korrekte Abbildung der Wirklichkeit bieten.

Üblicherweise steht am Beginn einer Krimiserie ein Verbrechen, welches das normale soziale Umfeld durcheinanderwirbelt. Für Gerechtigkeit zu sorgen und somit das Gleichgewicht wiederherzustellen, ist im Folgenden die Aufgabe der Hauptdarsteller von Krimiserien, die eine Erklärung sowie die für das Verbrechen verantwortliche Person finden und unschädlich machen müssen. In diesem Sinn erfüllen (die meisten) Krimiserien einen gesellschaftlichen Zweck, sie repräsentieren das Bild einer Gesellschaft, die auf der Basis von Recht und Ordnung funktioniert und in Balance bleibt. Der Hauptdarsteller genügt fast immer besonders hohen moralischen Ansprüchen – was ihn für das Publikum scheinbar über dem Gesetz stehen lässt und zu einer stärkeren Identifikationsfigur macht. Dass die Gesetzeshüter sich bei der Jagd auf Verbrecher dennoch streng an die Regeln der Legislative halten müssen, führt in Krimiserien oftmals zu Krisen.[209]

2.7.5. Dekodierung medialer Inhalte

Das Verstehen wird in den audiovisuellen Medien vor allem durch das Rezeptionstempo beeinflusst, das der Zuschauer nicht verändern kann – im Gegensatz zu visuellen Medieninhalten, etwa einem Buch oder eine Website. Die Aufmerksamkeit, die der Zuschauer einem Inhalt zukommen lassen kann, ist beschränkt, da er ansonsten den Anschluss an das Dargebotene verliert.[210]

Audiovisuelle Medien zeichnet aus, dass sie ihre Inhalte von den Rezipienten anhand verschiedener Möglichkeiten in einen Bedeutungszusammenhang zu ihrem eigenen Leben gestellt und somit verstanden werden können. Abgebildeten Szenen in Film und Fernsehen

[206] Vgl. Hickethier (2005), S. 14 ff.
[207] Vgl Faulstich (2008), S. 109
[208] Vgl. Burton (2000), S. 209
[209] Vgl. Bignell (2008), S. 130
[210] Vgl. Schlimbach (2007), S. 194

werden als soziale Handlungen begriffen, sowohl anhand von Bild und Ton als auch des narrativen Kontextes. Unterstützend werden gewisse Bereiche quasi leer gelassen, etwa durch abstrakte Bildsprache oder asynchronen Sound – der Zuschauer kann und muss den so entstandenen Leerraum selbst interpretieren. Dass der Mensch diese Leere mit Hilfe eigener kognitiver Leistung füllt, ist in Film und Fernsehen zwingend erforderlich – etwa wenn Ortswechsel und zeitliche Sprünge eingebaut werden.[211] Zum Beispiel sprechen zwei Detektive im Büro über ihren nächsten Einsatzort, die nächste Szene beginnt mit ihrem Ausstieg aus dem Auto. Dem Zuschauer wird nicht jede einzelne ihrer Handlungen (wie die Detektive ihre Mäntel anziehen, Autoschlüssel suchen, Büro absperren, auf die Straße gehen, ins Auto steigen, die Fahrt selbst, …) gezeigt – er erkennt aus dem Zusammenhang, dass die Detektive mit dem Auto von ihrem Büro aus zum Tatort gefahren sind. Leerräume sind also eine Grundvoraussetzung für Film und (insbesondere, da die Produktionen meist kürzer sind) Fernsehen, sie zu füllen ist eine der Verständnisleistungen der Zuschauer.
Der Interpretationsspielraum in den audiovisuellen Medien ist auch Teil der sozialen Kommunikation, wo innerhalb einer Familie oder dem Freundeskreis der Bedeutungsinhalt eines Films oder einer Fernsehproduktion diskutiert wird. Deutliche Verweise auf die Intentionen eines Film- oder Fernsehinhalts können allerdings ebenfalls Kommunikation entstehen lassen – wenn zum Beispiel ein Horrorfilm unter Fans des Genres berühmte Szenen nachstellt oder ein Protagonist ein Zitat aus einem Kultklassiker fallen lässt.

Um den Inhalt eines fiktiven Films oder Fernsehinhalts zu verstehen, sind laut Ohler[212] drei verschiedene Arten von Wissen nötig – narratives Wissen, generelles Weltwissen sowie das Wissen um filmische Darbietungsformen. Mit Hilfe des narrativen Wissens werden die Abläufe in bestimmten Genres erklärbar, das Verhalten der Protagonisten sowie typische Geschichten. So ist den meisten Menschen wahrscheinlich klar, dass eine bestimmte Tat (etwa ein Diebstahl im Krimi) eine Konsequenz nach sich zieht (Verfolgung, Festnahme, …), dass Quizshows üblicherweise ähnlich aufgebaut sind und wie ein Fußballspiel abläuft.
Das generelle Weltwissen ist weniger spezifisch – es hat jedoch eine wesentliche Aufgabe im audiovisuellen Medium: Es hilft dabei, die unvermeidbaren Lücken zu schließen (wie im Beispiel mit der Autofahrt oben). Ähnlich wie das narrative Wissen ermöglicht das Wissen um filmische Darstellungsformen das Verstehen der audiovisuellen Erzählung, Ohler[213] spricht von der Kombination als „narratives Form-Inhalt-Korrespondenzgitter“. Gewisse

[211] Vgl. Mikos (2008), S. 26 f.
[212] Vgl. Ohler (1994), S. 34 ff.
[213] Vgl. Ebenda, S. 37

ästhetische und auditive Darstellungsformen haben sich als Hinweise auf narrative Kontexte sowie emotionale Inhalte etabliert – zum Beispiel bestimmte Toneffekte, Kameraperspektiven oder Schnitte. Der Übergang zu einer nahen Kameraeinstellung, die lediglich das Gesicht des Protagonisten zeigt, steht so üblicherweise für Intimität und Nähe, die Froschperspektive hingegen vermittelt Größe oder sogar Übermacht und wird oft in Zusammenhang mit einer Bedrohung eingesetzt.

2.7.6. Empathie und Involvierung

Bei der Rezeption von audiovisuellen Medien können besonders starke Emotionen ausgelöst werden, da hier in kürzerer Zeit mehr und intensivere Ereignisse auftreten als im normalen Alltag des Zuschauers. Abgesehen davon werden diverse filmische Mittel eingesetzt, um eine stärkere Involvierung zu erreichen – schnelle Schnitte, Musik uvm. Der Zuschauer hat außerdem jederzeit die Möglichkeit, das emotionale Erlebnis abzubrechen – was diesem den Vorteil der Freiwilligkeit des Empfindens einbringt. Im echten Leben ist diese – besonders bei negativen Vorkommnissen – kaum gegeben. Nicht nur rezipiert der Zuschauer freiwillig, und damit auch die Emotionen, er sucht die audiovisuelle Stimulation aktiv, was die Emotionskomponente ebenfalls positiv steigert.[214]

Empathie kann ein Zuschauer diversen Untersuchungen anhand literarischer Texte zufolge vor allem dann erleben, wenn er in die fiktionale Welt eintauchen und sich auf diese konzentrieren kann – das gilt auch für audiovisuelle Medien. Der Zuschauer lebt sich in die Handlung hinein und nimmt den Platz der handelnden Person ein – etwa indem er überlegt, wie diese sich in der konkreten Situation verhalten wird und ob diese Entscheidung richtig ist. Es gibt verschiedene theoretische Überlegungen zu den Bedingungen für diese innere Beteiligung – ob der Zuschauer die filmische Welt für kurze Zeit als Realität ansieht (azentrale Imagination) oder konkret nachempfinden kann, was in filmischen Figuren vorgeht (zentrale Imagination), letzteres gewissermaßen die „Was wäre, wenn ich in dieser Situation wäre“-Erfahrung.[215] Ebenso umstritten wie die Art des Hineinlebens sind die konkreten filmischen Werkzeuge, mit denen Empathie beim Zuschauer ausgelöst wird.
Mikunda[216] spricht von „induzierter Spannung“ wenn visuelle Reize innere Erregung beim Zuschauer auslösen. Diese Erregung kann durch emotionale Bildkomposition, etwa unterschiedlich starke Spannungspunkte, das Komprimieren oder die Ausdehnung des

[214] Vgl. Mangold, in: Bartsch / Eder / Fahlenbrach (Hg.) (2007): S. 175 f.
[215] Vgl. Bruun Vaage, in: Schick / Ebbrecht (2008): S. 30 ff.
[216] Vgl. Mikunda (2002), S. 34 ff.

visuellen Feldes, aber auch das Verleihen von visuellem Gewicht (etwa durch Schärfen eines Objekts, das aus der Masse heraussticht) hervorgerufen werden. Schräge Kameraansichten können ebenso zu einer stärkeren Involvierung führen wie die Verformung von Objekten und Ansichten. Außerdem spielt der Reizwechsel eine große Rolle dabei, ob uns ein Film oder eine TV-Produktion gefangen nehmen können – das Auge verlangt nach Bewegung, das Ohr wird von Stimme und alarmierenden Geräuschen aktiviert und eingebunden. Große Bedeutung beim Hineinziehen des Zuschauers kommt auch der Musik zu, da Rhythmus das natürliche Bedürfnis weckt, mitgerissen zu werden

2.7.7. Spezifika des Fernsehens

Das Medium Fernsehen wird von einigen Eigenheiten bestimmt, die nicht auf den „großen Bruder“ Kino zutreffen. Burton[217] beschreibt diese als die Freude am Wiedererkennen, der Familiarität des Mediums. Die wiederholte, oft sogar tägliche Hinwendung zum Medium Fernsehen gibt Sicherheit, außerdem sorgt die Wiederholung von Material, Genres und Serien für Befriedigung. Die Ähnlichkeit der Erfahrung durch das Fernsehen sorgt außerdem für Gesprächsstoff und stärkt so soziale Bindungen – Fernsehen hat sich zu einer gemeinschaftlichen sozialen und kulturellen Erfahrung entwickelt. Zuseher können sich besonders bei Serien oder seriellen Formaten (wie Quizshows) außerdem darauf verlassen, involviert zu sein – indem sie zum Beispiel versuchen, den Schuldigen in einer Krimiserie zu erkennen. Diese Involvierung führt zu einem Versinken im Gezeigten, der Fernsehzuschauer kann so Erfahrungen machen ohne Verantwortung inne zu haben.
Hickethier[218] beschreibt das Fernsehen als permanente und im Vergleich zum Kino wesentlich mächtigere Emotionsmaschinerie, da zum Beispiel die Deutschen ab dem Alter von 3 Jahren pro Tag etwa dreidreiviertel Stunden mit Fernsehen beschäftigt sind – aber nur vier Stunden pro Jahr im Kino verbringen. Das Bedürfnis nach medial vermittelten Emotionen hängt unter anderem mit dem Verlangen nach gesellschaftlicher Einheitlichkeit und Sozialisation zusammen. Außerdem können über das Fernsehen als Mittler Emotionen ausgelebt werden. Etwa wenn ein großer Teil der arbeitslosen Bevölkerung fast pausenlos Talk- und Gerichtsshows rezipiert, die Neid, Hass und Zorn regelrecht zelebrieren – und so vormachen, dass es Menschen gibt, die noch weniger Glück im Leben haben.
Fernsehen ist weniger von Stars per Kino-Definition abhängig sondern mehr von Persönlichkeiten, denen von den Zuschauern bestimmte Qualitäten zugeschrieben werden.

[217] Vgl. Burton (2000), S. 71 f.
[218] Vgl. Hickethier, in: Bartsch / Eder / Fahlenbrach (Hg.) (2007): S. 118

Durch die nur gelegentliche Berührung mit dem Kino entsteht zu dessen Stars laut Burton[219] eine Distanz, die ihnen gleichzeitig fast mystische Erhöhung zukommen lässt. Über das Fernsehen, die regelmäßige und oftmalige Hinwendung zum Medium, hingegen bauen die Zuschauer eine intimere Beziehung zu den Schauspielern auf. Dieses Kreieren von Persönlichkeiten ist ein Muss für die meisten Fernsehproduktionen – seien es Serien, Nachrichten oder Shows aller Art.

2.8. Sound im audiovisuellen Medium

Mit Einführung des High-Fidelity-Standard in den 1960er Jahren (zur geschichtlichen Entwicklung siehe auch 2.9.1), war es erstmals möglich, Töne klanggetreu wiederzugeben – die Zeit des Rauschens, verzerrter Stimmen und ähnlicher technischer Unzulänglichkeiten näherte sich dem Ende. Aufzeichnung und Wiedergabe wurden einander angeglichen – was jedoch ein neues Problem mit sich gebracht hat: Der Mensch ist in seiner natürlichen Lebensumgebung ständig von Geräuschen umgeben, wenn auch oft nur sehr leise und im Hintergrund. Um das Fehlen dieser natürlichen Soundkulisse wett zu machen, entstanden neue Techniken und Betätigungsfelder, etwa das nachträgliche Hinzufügen von Geräuschen wie unverständlichem Gemurmel oder Straßenlärm. Damit haben audiovisuelle Medien und das Radio eine eigene auditive Wirklichkeit geschaffen, die meist nichts mit dem Klangteppich der „natürlichen“ menschlichen Umgebung abseits der Medien zu tun hat.[220]

Die Medien haben sich im auditiven anders als im visuellen Feld der menschlichen Gewohnheit angeglichen. So ist etwa für das Konsumieren eines Kinofilms eine Veränderung der natürlichen Sehgewohnheit vonnöten – schnelle Schnitte, unnatürliche Geschwindigkeitswechsel wie Zeitlupeneffekte oder das Fehlen einer Tiefe im Raum kommen in der natürlichen Umgebung des Menschen nicht vor. Er passt sein Sehverhalten und das Verständnis für das Gesehene also dem Medium an – im Gegensatz zum Gehör.

Dennoch kommt die menschliche Wahrnehmung den audiovisuellen Medien entgegen, denn obwohl Geräuschquellen nicht aus der Richtung wahrgenommen werden, aus der sie eigentlich stammen müssten – sprich, nicht direkt von einer bestimmten Stelle des Bildschirm bzw. der Kinoleinwand – werden sie als dem Bild zugehörig akzeptiert. Dieses Zusammenfügen wird als „Bauchredner-Effekt“ oder „Visual Capture“ bezeichnet. Begründet ist diese Art der Wahrnehmung in der natürlichen Umwelt, wo ein Schallereignis

[219] Vgl. Burton (2000), S 77 f.
[220] Vgl. Westphal (2002), S.157

normalerweise von jenem Ort stammt, an dem wir eine zugehörige Schallquelle sehen.[221] Auch hier sind Zusammenhänge mit der Gestaltpsychologie zu erkennen (siehe 2.8.3)
Im Gegensatz zu den Bildern in audiovisuellen Medien, die einzeln aufgezeichnet werden, muss Ton kontinuierlich aufgenommen werden um rezipiert werden zu können. Es gibt wegen der unterschiedlichen Wahrnehmung von Augen und Ohren auch keine tonale Entsprechung für das Standbild – Ton existiert nur auf einer Zeitlinie. Er kann außerdem ohne einer grundsätzlichen Veränderung zu unterliegen nicht verkürzt oder verlängert werden.[222]

Chion[223] zufolge gibt es beim Film keinen „Soundtrack" an sich, es gibt lediglich einen Ort für Bilder und Ton. Synchron-Sounds, die im Zusammenhang mit dem Bild wahrgenommen und von diesem in der menschlichen Wahrnehmung „verschluckt" werden, beweisen Chion zufolge, dass es keinen eigentlichen „Soundtrack" gibt – da das Bild ohne den Ton, oder wenn dieser verändert wurde, nicht mehr dasselbe ist.
Der Sound besteht zwar aus verschiedenen Einheiten – Sprache, Geräusche, Musik –, diese entsprechen jedoch der Hörerfahrung, die wir auch im Alltag machen und sind nicht künstlich hergestellt wie der Schnitt des Bildes auf verschiedene Personen, Gebäude oder Gegenstände, der nicht der natürlichen Wahrnehmung des menschlichen Auges entspricht. Entsprechend werden die Soundeinheiten vom Zuschauer auch nach alltäglichen Kriterien beurteilt und eingeteilt – nicht nach spezifisch filmischen. Die auditiven Elemente eines Films werden nicht autonom rezipiert – sie werden in einen Zusammenhang gesetzt zum Gesehenen, Chion spricht von einer „unmittelbaren Selektierung durch die Wahrnehmung"[224]. Auditive Elemente werden also sofort in einen Zusammenhang gesetzt zu den visuellen– sie können quasi vom Bild „geschluckt" werden, wenn sie synchron laufen (siehe 2.8.1.2) oder Aufmerksamkeit erregen, wenn sie aus dem Off kommen. Außerdem kann Sound vom Zuseher in einen imaginären Raum, ähnlich einer Vorbühne im Theater, versetzt werden – das trifft zum Beispiel auf Hintergrundmusik oder einem Erzähler aus dem Off zu.

2.8.1. Aufgabe des Tons

Mit Hilfe von Ton können dem Zuschauer im audiovisuellen Medium die Dimensionen Zeit und Raum vermittelt werden, da er allgegenwärtig ist und – vor allem anhand der Effekte und Geräusche, des Soundtracks – Atmosphäre sowie Kontinuität schafft.[225]

---

[221] Vgl. Goldstein / Irtel (Hg.) (2008), S. 310 f.
[222] Vgl. Monaco (2000), S. 122
[223] Vgl. Chion (1994), S. 40 ff.
[224] Vgl. Chion (1999), S. 3
[225] Vgl. Monaco (2000), S. 215

Zettl[226] zufolge hat Sound diverse Funktionen: er liefert verbale Informationen – durch Dialog, direktes Ansprechen und Erzählung – und dient der äußeren Orientierung in Bezug auf den filmischen Raum und seinen Schauplatz, die Zeit und spezielle Situation (etwa das Weinen eines Babys). Außerdem vermittelt Sound Hinweise auf die innere Orientierung – wie die Stimmung (meist durch Musik vermittelt) und das Befinden einer Person (etwa schnelles Atmen als Zeichen für Aufregung) –, die Energie und Strukturierung einer Szene (beides ebenfalls meist durch Musik vermittelt).
Sound hat laut Chion[227] vor allem den Zweck, Bilder zu verbinden und Atmosphäre herzustellen, die das Bild unterstützt. Außerdem kann Sound in einer Szene die Funktion als Zeichensetzer übernehmen – durch Stille, ein plötzliches Geräusch oder Musik. Mit Hilfe von Sound ist es zudem möglich, eine Erwartungshaltung beim Zuseher auslösen, die entweder erfüllt oder überraschend aufgelöst wird – der Sound erfüllt also eine konvergente oder divergente Aufgabe.

2.8.1.1. Zusatznutzen

Chion[228] bezeichnet die Informationen, die die Soundkulisse im Film einem Bild hinzufügt, als „Added Value" – das Gezeigte wird verstärkt, es kann allerdings auch eine Diskrepanz dazu aufgebaut werden oder kein direkter Zusammenhang zum Gesehenen bestehen. Chion widerspricht der Meinung, das Bild würde auch ohne Ton die gleichen Informationen oder Eindrücke vermitteln. Ton gibt dem Bild eine (möglicherweise völlig neue) Struktur – er rahmt es quasi ein.[229] Da das Ohr Sinneseindrücke schneller verarbeitet als das Auge, kann der Ton unterstützend wirken, um mögliche Überforderung oder Verwirrung durch visuelle Eindrücke zu vermeiden – zum Beispiel durch den Einsatz von Rufen oder Soundeffekten als Markierung bei schnellen Bewegungen in Kampfsport-Szenen in Filmen.[230]
Ob der Zuschauer den „Added Value" des Tons annimmt, der dem Bild möglicherweise eine völlig neue Bedeutung gibt, hängt vor allem davon ab, ob der Sound als synchron und plausibel empfindet. So kann dasselbe Geräusch in einer Komödie als genauso passend empfunden werden wie in einem Kriegsfilm.[231]
Besondere Bedeutung hat der Zusatznutzen des Tons Chion zufolge dank des „Synchresis"-Prinzips.

---

[226] Vgl. Zettl (1995), S. 341 ff.
[227] Vgl. Ebenda, S. 48 ff.
[228] Vgl. Ebenda, S. 5
[229] Vgl. Ebenda, S. 7
[230] Vgl. Chion (1994), S. 10 f.
[231] Vgl. Ebenda, S. 22 f.

2.8.1.2.„Synchresis“ und Synchronizität

Der Begriff „Synchresis“ wurde von Chion[232] geprägt, der Synchronizität und Synthese mischte. Unter Synchresis versteht er das unwillkürliche, unterbewusste Verlangen des Menschen, Gehörtes mit Gesehenem zu verbinden, wenn die Ereignisse zur selben Zeit auftreten – Synchresis ist somit die Grundlage für das Gelingen von Synchronisation, Sound-Effekten und der Tonmischung. Synchresis passiert trotz der Unbewusstheit nicht automatisch – das Prinzip basiert etwa auf richtiger Bedeutung, Rhythmus sowie Regeln der Gestaltpsychologie (siehe 2.8.3).[233] Dazu zählt zum Beispiel die Tendenz, ähnliche oder eng beieinander liegende Elemente zusammenzufassen und als geschlossene Einheit zu betrachten.

Am „Point of Synchronization“ treffen sich Bild und Sound in Synchronizität, das kann eine Veränderung, Akzentuierung oder auch den Bruch des audiovisuellen Flows bedeuten, aber auch eine Sequenz stimmig abschließen sowie einem Bild semantischen Charakter verleihen – etwa indem sich ein stark betontes Wort mit einem ebenso stark aufgeladenen Bild vereint. Nicht immer muss der Sound dabei einem in der Realität vertrauten entsprechen – beim „Emblematic Synch Point“, also dem symbolischen Punkt der Synchronizität, kommt zum Tragen, dass ein Geräusch ein Bild unterstützt oder verstärkt, das alleine zu schwach wäre um die Bedeutung zu transportieren, die es innehat. Bekanntestes Beispiel: Die Verstärkung von dargestellten Schlägen durch comichafte „Punch“-Geräusche. Synchronizität ist verschieden stark ausgeprägt, vor allem in Bezug auf Lippensynchronizität in verschiedenen Sprachen.[234]

2.8.2. Unterteilungskriterien

Man kann Sound im audiovisuellen Medium nach verschiedenen Kriterien einteilen – Zettl[235] schlägt die Unterteilung in „literal“ und „non-literal“ vor, übersetzt bedeutet das etwa „wörtlich“, bzw. „buchstäblich“ sowie das Gegenteil. Stimme zählt zu den Literal Sounds, da sie Bedeutung transportiert und – auch wenn der Sprecher nicht im Bild ist („asynchroner Sound“) – Rückschlüsse darauf zulässt, was (also ein Mensch) oder wer (eine bestimmte Person, die bereits im Bild war) den Sound produziert. Non-literal Sound, auch „non-diegetic“ genannt, ist meist Musik oder auch Geräusche in Cartoons.

---

[232] Vgl. Gregory (2001), S. 16 f.
[233] Vgl. Chion (1994), S. 63
[234] Vgl. Ebenda, S. 58 ff.
[235] Vgl. Zettl (1995), S. 338

Metz[236] hat die Unterscheidung zwischen „diegetic“ und „non-diegetic“ – also „erzählend“, „erörternd“ bzw. das Gegenteil – für das gesprochene Wort im Kino eingeführt (siehe 2.9.7). Darauf aufbauend gibt es zahlreiche Unterscheidungen für Sound im Film allgemein. Sonnenschein[237] findet verschiedene, grundlegende Einteilungen für Sound – neben „diegetic“ und „non-diegetic“ unterscheidet er in „On Screen“- und „Off Screen“-Sound, je nachdem, ob die Geräuschquelle gerade im Bild sichtbar ist oder nicht. Unterteilen kann man Sonnenschein zufolge außerdem in „aktiven“ und „passiven“ Sound. Während ersterer neugierig macht und Fragen aufwirft, stellt passiver Sound Atmosphäre her.

Percheron unterteilt „marked“ und „unmarked Diegesis“, je nachdem, ob ein Film den Erzählakt in den Vordergrund stellt oder nicht. Eine andere Differenzierung nehmen David Bordwell und Kristin Thompson vor. „Simple diegetic Sound“ scheint von einer Quelle aus dem Bild hervorzugehen und ist zum Teil simultan zu sehen. „External diegetic Sound“ hat seinen Ursprung im Empfinden der Zuseher ebenfalls im Bild, die Charaktere sind sich dieses Sounds bewusst. Im Gegensatz dazu steht der „Internal diegetic Sound“, der der Vorstellung eines einzelnen Charakters entspringt und dementsprechend von anderen Charakteren nicht wahrgenommen wird. „Displaced diegetic Sound“ entspringt zwar dem Bild, der Ursprung liegt jedoch vor oder nach den gezeigten Bildern. „Non-diegetic Sound“ schließlich ist der Sammelbegriff für andere Sounds, etwa Hintergrundmusik ohne erkennbare Geräuschquelle oder einen unbekannten Erzähler.[238]

2.8.3. Gestaltungsregeln

Die Gestaltpsychologie nahm ihren Anfang im Deutschland der 1920er-Jahre, wo zuerst von subjektiven Berichten – was der Theorie Kritik einbrachte – abgeleitet wurde, dass der Mensch in seiner Wahrnehmung zur Geschlossenheit tendiert. So sieht der Mensch laut Gestaltpsychologie ein in etwa kreisförmiges Muster von Punkten dergestalt, als „gehöre“ es zu einem Objekt und bilde ein Objekt. Außerdem werden demnach Teile, die sich gemeinsam bewegen – zum Beispiel Blätter an einem Baum – als ein Objekt wahrgenommen. Trotz der Kritik durch die großteils unwissenschaftlichen Methoden gelten viele Gesetze der Gestaltpsychologie bis heute – im visuellen sowie im auditiven Bereich.[239]

---

[236] Vgl. Stam / Burgoyne / Flitterman-Lewis (1994), S. 60
[237] Vgl. Sonnenschein (2001), S. 153
[238] Vgl. Stam / Burgoyne / Flitterman-Lewis (1994), S.60 f.
[239] Vgl. Gregory (2001), S. 16 f.

Sonnenschein[240] baut einige Regeln für Sound im Film auf der Gestaltpsychologie auf. Diese Grundsätze lassen Sound im audiovisuellen Medium verständlich wirken und ermöglichen die Interpretation von Geräuschen sowie Zusammenhängen mit dem Bild. Während mehrere Geräusche in der Wahrnehmung meist zu einem verschwommenen Ganzen und damit auditiv schlecht auswertbar werden, kann eine einzelne Stimme – zum Beispiel der Ruf eines Bekannten in einem Restaurant – hervorstechen (siehe 2.4.4.3). Ebenfalls auf der Gestaltpsychologie aufbauend, stellt Sonnenschein fest, dass das Gehirn vollständige Muster, die Ganzheit, bevorzugt. Werden Sound im Allgemeinen oder Stimme im Speziellen abgeschnitten oder aufgehalten, bevor sie das erwartete Ende erreichen, wirkt sich dies negativ auf den Menschen aus – Gefühle des Konflikts und der Spannung sind die Folge. Unter „Good Continuation" versteht Sonnenschein, dass ein Soundeffekt trotz Änderungen – zum Beispiel der Intensität, des Ortes oder der Frequenz – weiterhin als fortgesetztes Geräusch einer einzigen Quelle erkannt wird. Abrupte Umbrüche hingegen lassen das Gehirn von einer neuen Quelle ausgehen. Die menschliche Stimme ist von sanften Frequenzänderungen geprägt, die Äußerungen als von einer einzigen Person kommend identifizierbar machen.

Unabhängig davon versucht das Gehirn, lose Enden zu verknüpfen, indem es zwei eigentlich nicht verbundene Teile zu einem Ganzen zusammenfügt – es ist auf den Abschluss aus. Dieser Grundsatz ermöglicht im audiovisuellen Medium zum Beispiel, dass ein Autohupen die Lücke in einem fehlerhaften Dialog schließt – das Publikum nimmt keine Unterbrechung wahr, da das Gehirn die beiden Teile des Gesprächs als zusammengehörig einstuft und verbindet. Die Erkenntnis aus der Gestaltpsychologie, dass das Gehirn nahe beieinander liegende Objekte zu einem Ganzen gruppiert, kann neben dem Sehen ebenfalls auf das Gehör übertragen werden. So werden verschiedene Geräusche als zusammengehörig wahrgenommen, wenn sie zeitnah hintereinander zu hören sind –im Film kann ein Quietschen, dem ein Schlag und ein Knarren folgen, für das Geräusch einer rostigen Tür stehen und somit vom Gehör als ein zusammengehöriges Geräusch wahrgenommen werden. Unverständliches Murmeln einer Menschenmenge oder Klatschen sind weitere typische Beispiele für das Gruppieren von Gehörtem.

Unabhängig vom Faktor Nähe spielt auch jener der Ähnlichkeit eine große Rolle beim Höreindruck. So werden ähnliche Sounds, auch wenn sie zeitlich getrennt sind, als

[240] Vgl. Sonnenschein (2001), S. 79

zusammengehörig wahrgenommen – etwas das von Bürolärm immer wieder unterbrochene Klingeln eines Telefons. Als „Common Fate“ bezeichnet Sonnenschein jenes Phänomen, wenn verschiedene Teile eines vielschichtigen Sounds den gleichen Veränderungen zur selben Zeit unterliegen und daher als von einer Quelle stammend wahrgenommen werden. Diese Regel findet im audiovisuellen Medium zum Beispiel Anwendung, wenn Hupen, Autolärm und Gemurmel beim Wechsel von einem Außenschauplatze nach drinnen leiser werden, der Dialog der Figuren im Bild aber gleich laut bleibt. Der Zuseher setzt Hupen, Autolärm und Gemurmel zu einem gemeinsamen Höreindruck als Straßenlärm zusammen. Eine weitere Regel ist die der „Zugehörigkeit“, darunter versteht Sonnenschein, dass ein bestimmter Teil des Sounds nur als von einer einzigen Quelle stammend wahrgenommen wird und nicht ohne weiteres auf eine andere Quelle übertragen werden kann. Passiert dies jedoch geschickt, kann diese Übertragung ein subtiles dramaturgisches Mittel darstellen.[241] Das könnte zum Beispiel der Fall sein, wenn zuerst ein tropfender Wasserhahn samt Geräusch im Bild ist, die Kamera schwenkt und nur mehr das tropfende Geräusch zu hören, der Wasserhahn aber nicht mehr zu sehen ist – bis Bluttropfen im Bild auftauchen. Der Höreindruck verschiebt sich also von der Quelle Wasserhahn zum Blut – eine audiovisuelle Überraschung.

2.8.4. Sound im Fernsehen

Anders als das Kino war das Fernsehen von Anfang an ein audiovisuelles Medium, und anders als das Kino braucht das Fernsehen Sound für die Kommunikation, wie Zettl[242] postuliert. Fernsehen ist mehr als Kino die Reflektion von Realität, ob es sich nun um Dokumentationen, Nachrichten oder Krimiserien handelt. Der Soundtrack muss also den Bildern Glaubwürdigkeit verleihen, nicht umgekehrt – was dem Sound eine größere Bedeutung gibt. Im Fernsehen ist lediglich eine wesentlich kleinere Auflösung als im Kino möglich, was sowohl auf die visuelle Gestaltung als auch den zugehörigen Soundtrack Auswirkungen hat – so fesseln etwa hochaufgelöste Landschaftsbilder im Kino ohne Sprache wesentlich länger die Aufmerksamkeit des Zusehers als die niedrig aufgelöste Version im Fernsehen. Durch die schlechtere Auflösung und den kleineren Bildschirm ist es im Fernsehen außerdem schwieriger, Informationen nur über Bilder zu transportieren – der Sound muss zusätzliche Details liefern

[241] Vgl. Sonnenschein (2001), S. 80 ff.
[242] Vgl. Zettl (1995), S. 334 f.

2.9. Stimme im audiovisuellen Medium

Chion[243] bezeichnet den Ton im Kino als vokozentristisch, die Stimme steht also im Mittelpunkt. Schon während der Filmaufnahmen geht es bei den zugehörigen Tonaufnahmen vor allem um die Aufzeichnung der Stimmen, die von der übrigen Geräuschkulisse isoliert werden. Schließlich nimmt der Mensch zuallererst eine Stimme wahr, bevor er der umgebenden Geräuschkulisse wie Wind oder Musik seine Aufmerksamkeit schenkt.
Stimme kann im audiovisuellen Medium abgesehen von den gesprochen Inhalten auch durch ihr Wesen an sich Hinweisgeber sein. Geschlecht, Akzent, Lautstärke, Timbre und Tonfall sowie Sprechgeschwindigkeit wecken im Zuschauer Stimmungen. Außerdem lässt jener Ort, von dem aus die Stimme zu kommen scheint – innerhalb der Erzählung oder aus dem Off – Rückschlüsse auf die Art des Gesehenen sowie die Erzählstruktur zu.[244]

Stimme im Medium unterscheidet sich grundlegend von der natürlichen Stimme - sie ist meist aufgezeichnet, oft geschnitten oder technisch verändert und entspricht damit nicht mehr der natürlichen Sprachgewohnheit und der Gebundenheit an eine Person, einen Ort und einen Zeitpunkt. Westphal[245] stellt außerdem fest, dass durch das Play-back-Verfahren eine „Ideal"-Stimme geschaffen werden kann, die von natürlichen Gegebenheiten wie der Gesundheit eines Menschen oder dem Altern abgekoppelt existiert.
Nicht nur technisch unterscheidet sich Stimme im Medium vom üblichen natürlichen Höreindruck. Sonnenschein[246] konstatiert, dass es für psychologisch gesunde Menschen meist nicht nötig ist, ihre Emotionen in der Stimme zu verstecken – diese werden stattdessen im Einklang mit dem Gesagten vermittelt. Subtext und gesprochene Wörter verbinden sich also üblicherweise zu einem gemeinsamen Höreindruck. Bei Film und Fernsehen existiert diese Verbindung jedoch oftmals nicht, so Sonnenschein:"(…) and it is exactly the subtleties and conflicts between the obvious and hidden agendas that make for good drama.".[247]

Trotz oder vielleicht gerade wegen der Unangepasstheit des Gehörs versucht der Mensch, Geräusche und ganz besonders Stimme auch in medialen Situationen in einen Bezugsrahmen zu setzen. Da Geräusche insbesondere der Lokalisierung dienen, ist der Mensch erpicht darauf, diese Funktion trotz der technischen Übermittlung von Stimme und Sprache zu erhalten. Westphal skizziert diese Suche nach Ersatz für fehlende auditive Informationen mit

[243] Vgl. Chion (1994), S. 5 f.
[244] Vgl Bignell (2008), S. 157
[245] Vgl. Westphal (2002), S. 157
[246] Vgl. Sonnenschein (2001), S. 146
[247] Vgl. Ebenda, S. 146 f.

der üblichen Frage bei einem Handygespräch: „Wo bist du?“[248] So kompensiert der Mensch die fehlende Orientierung, die moderne Technik aufgrund der Verschiebung von Ort – und, im Fall anderer Medien, auch Zeit – mit sich bringt.

2.9.1. Geschichte der Stimme im Medium

Um hörbar zu sein, musste Stimme zu Beginn der Aufzeichnungsgeschichte mit dem Phonographen, dem Edison zum Durchbruch verhalf, regelrecht geschrien werden – was naturgemäß sowohl den Sprecher und seine Stimmeigenschaften verzerrte als auch das Erleben der Zuhörer. Als1925 Aufnahmen mit Grammophonen möglich waren, die von Elektromotoren betrieben wurden, besserte sich die Stimmqualität, da nun die Aufnahmen mittels Mikrofon verstärkt werden konnten. Störgeräusche und Verzerrungen gehörten allerdings weiterhin zur medial vermittelten Stimme. Die Nationalsozialisten verhalfen 1942 dem Tonband zum Durchbruch, das den Grundstein für die moderne mediale Erfahrung von Stimme legte: Erstmals gab es damit die Möglichkeit, zu schneiden und Übergänge fließend zu gestalten. Wirklich naturgetreue Sprachwiedergabe ermöglichte allerdings erst High Fidelity in den 1960er Jahren.[249] Unzählige neue Arten, Stimme zu verändern, ermöglichte die Digitalisierung des späten 20. Jahrhunderts – berühmt-berüchtigt sind etwa geschönte Gesangsaufnahmen von Sängerinnen und Sängern, die bei Live-Auftritten keineswegs mit stimmlichen Qualitäten punkten können. Westphal[250] konstatiert, dass durch die Harmonisierung und Bearbeitung der Stimme in den Medien die eigene Stimme als minderwertig empfunden wird, sie spricht von einer „ästhetischen Normierung“ durch die Medien.

2.9.2. Geschichte der deutschsprachigen Synchronisation

Als sich in den 1930er Jahren der Tonfilm durchsetzte, konkurrierten in den ersten Jahren drei Übersetzungstechniken für fremdsprachige Filme: Neben Untertiteln wurden Filme vor allem zwischen 1929 und 1932 mehrsprachig – und damit sehr kostenintensiv – gedreht, danach setzte sich die technisch und qualitativ noch unausgereifte Synchronisation durch.[251]
Durch die Anpassungen an die deutsche Mentalität kam von erstaunten Reaktionen von Seiten der Kritiker bis hin zu Vorwürfen des Betrugs und der Fälschung. Für Gegner des erstarkenden nationalsozialistischen Regimes, denen der Auftritt in Filmen und auf der Bühne

[248] Vgl. Westphal (2002), S. 132
[249] Vgl. Ebenda, S. 156 f.
[250] Vgl. Westphal (2002), S. 158
[251] Vgl. Bräutigam (2001), S. 9 f.

bereits verboten war, bot sich mit der Synchronisation bis zum Jahr 1941 eine Ausweichmöglichkeit. Ab dann waren allerdings nur mehr wenige ausländische, zum Beispiel einige italienische und spanische, Filme erlaubt – US-amerikanische Filme waren in Deutschland und Österreich ab diesem Jahr generell verboten. Nach Kriegsende begann die Blütezeit der Film-Synchronisation, da die US-amerikanischen Besatzer ihre Filme in Deutschland und Österreich zur „Reeducation" einsetzten und sich die so lange vermissten ausländischen Filme in den Kinos großer Beliebtheit erfreuten. Die Synchronisation trug ihren Teil als vermittelndes Element zwischen den im Krieg unterlegenen Länder und ihren Besatzern bei. Nach dem Krieg wurden daher bereits vor dem Entstehen privater Synchron-Ateliers diverse Synchronisationsstudios von den Besatzern in verschiedenen Teilen Deutschlands eröffnet.[252]

2.9.3. Abläufe der Synchronisation

Vor der eigentlichen Synchronisation steht heute naturgemäß eine professionelle Rohübersetzung des gesprochenen Textes, die anschließend von einem Synchronautor bearbeitet wird. Er muss den Inhalt passend zu den im Bild sichtbaren Mundbewegungen und dem Schnitt umsetzen. Vor allem Labiallaute (im Deutschen p, k, t) sollten bei Großaufnahmen synchron gesprochen werden, ansonsten wird die Illusion gestört – alle Silben exakt den Mundbewegungen anzugleichen wird seit den 50er Jahren jedoch nicht mehr als nötig erachtet.

Nach der Bearbeitung des Textes durch den Synchronautor ist der Synchron-Regisseur am Zug, der die Synchronsprecher in ihre Rollen einweist und ihnen Anleitungen für Ausdruck und Modulation gibt. Schließlich sind die Sprecher mit einzelnen Takes konfrontiert, das sind einige Sekunden kurze Sequenzen, in die der Film zerlegt ist – sie kennen meist weder den Film noch die handelnden Personen oder den kompletten Dialog. Der Cutter überwacht, dass die Stimmaufnahmen synchron zum Bild laufen, anschließend passt er die Aussagen im Schneideraum an das Bild an. Akustische Effekte wie Echo oder Telefonstimmen werden zuvor noch vom Tonmeister eingefügt. Letzte Hand an die Synchronisation wird mit der Tonmischung angelegt, wo die aufgenommenen Synchronstimmen mit dem Soundtrack des Original – dem sogenannten „IT-Band" (für: International Tape) – abgemischt werden. Auf dem IT-Band sind alle Geräusche der Originalaufnahmen abgesehen von den Stimmaufnahmen zu hören.[253]

---

[252] Vgl. Bräutigam (2001), S. 10 ff.
[253] Vgl. Ebenda, S. 34 f.

2.9.4. Die Synchronstimme

Im Gegensatz zur Arbeit als Schauspieler muss der Synchronsprecher auf die körperliche Arbeit, die Einheit von Sprache und Bewegung verzichten. Außerdem muss er abgesehen vom Inhalt auf unabhängige Faktoren wie das vorgegebene Sprechtempo und die Bilddynamik achten.[254]
Synchronstimmen haben einen enormen Wiedererkennungswert, auch wenn die Person dahinter völlig unbekannt ist und die Stimme möglicherweise mit mehreren Filmpersönlichkeiten in Zusammenhang gebracht wird. Über den höchsten Wiedererkennungswert verfügen Synchronstimmen, die in Fernsehserien zu hören sind, die regelmäßig in kurzen Abständen und über längere Zeit hinweg ausgestrahlt werden. Ein Wechseln der Synchronstimme für einen Akteur zerstört daher die aufgebaute Illusion und wirkt entfremdend. Dennoch besteht der „harte Kern“ der deutschsprachigen Synchronsprecher nur aus zwei bis drei Dutzend Menschen, die neben einem oder gar mehreren Star-Schauspielern zahlreiche weitere Hauptrollen vertonen – was das Publikum irritieren kann.[255]

Durch die Synchronstimme werden dem Schauspieler bestimmte Persönlichkeitsmerkmale zugeordnet. Bräutigam[256] zufolge ist daher weniger wichtig, ob die Originalstimme der Synchronstimme ähnelt, als wie die Synchronstimme zum Charakter in Film oder Serie passt. Gundermann[257] betont, dass es auffallend sei, „wie bei gelungener Übertragung Original- und Synchronsprecher sich einander ähneln.“ Mit der Entscheidung für eine bestimmte Synchronstimme wird dem Zuseher also gleichzeitig bereits eine Interpretation für die dargestellte Person geliefert, die im Original fehlt. Durch die akustische Dominanz kann der Charakter in der synchronisierten Fassung eine sich vom Original unterscheidende Stereotypisierung erfahren und auf den Zuseher sogar interessanter wirken.[258]

2.9.5. Klanggetreue Sprachwiedergabe

Bei klanggetreuer Sprachwiedergabe muss vor allem die Höhe der Formanten beachtet werden, die besonders bei Vokalen auftreten und Frequenzen von 64 bis 16.384 Hz umfassen.

---

[254] Vgl. Bräutigam (2001), S. 26
[255] Vgl. Ebenda, S. 27 f.
[256] Vgl. Ebenda, S. 29 f.
[257] Vgl. Gundermann (1994), S. 45
[258] Vgl. Bräutigam (2001), S. 29 f.

Bis zu einer Übertragung von 64 bis 4096 Hz bleibt die Sprache verständlich, von 64 bis 2048 Hz werden noch etwa 90 Prozent der Wörter richtig verstanden.[259]
Sonnenschein[260] beschreibt in seinem Buch „Sound Design“ die verschiedenen Möglichkeiten, Stimme im Film hörbar zu machen. Zu hohe oder zu niedrige Aufnahmelautstärke, störende Hintergrundgeräusche oder technische Unzulänglichkeiten – etwa bei der Aufzeichnung mehrerer Stimmen gleichzeitig – können eine Nachbearbeitung der Stimme im audiovisuellen Medium nötig machen. Möglich ist das zum Beispiel durch eine erneute Aufzeichnung der Dialoge am Set, möglichst zeitnah zum Drehen einer Szene – „Wild Track“ genannt. Eine Alternative dazu ist, die Tonspur einer Einstellung des Films über die Bildspur einer anderen, visuell besseren Einstellung der gleichen Szene zu legen. Diese Methode wird jedoch bevorzugt dann angewendet, wenn der sprechende Charakter nicht im Bild zu sehen ist, da die Lippensynchronizität ansonsten kaum gegeben ist. Die gebräuchlichste Arbeitsweise zur Nachbearbeitung von stimmlichen Äußerungen ist im Film jedoch das Postproduction Dubbing, die Nachvertonung mit dem Schauspieler im Studio.[261]

2.9.6. Vokozentrismus im Medium

Die Stimme spielt in der menschlichen Hörwahrnehmung eine ganz besonders wichtige Rolle (siehe 2.4.4.2), wodurch sie auch im audiovisuellen Medium einen speziellen Platz vor allen anderen Sounds einnimmt. Um diesem Umstand gerecht zu werden, muss die Stimme im Medium künstlich hervorgehoben werden. Chion[262] zufolge ist dies vor allem deshalb nötig, weil im audiovisuellen Medium jene stereophonen Unterscheidungsmöglichkeiten fehlen, die die Stimme in der Natur von Umgebungsgeräuschen abheben. Das Abnehmen des Tons beim Dreh ist daher meist auf die Stimme fixiert (siehe 2.9.5).

2.9.7. Diegese

Metz[263] hat drei Kategorien für Stimme und Sprache im Film eingeführt: Bei der „fully diegetic speech“, also dem gänzlich diegetischen (erzählenden, erörternden) Sprechen, ist der Charakter zu sehen, der gerade spricht. Im Gegensatz dazu steht die „non-diegetic speech“, zum Beispiel ein unbekannter Erzähler aus dem Off. Eine Mischform stellt die „semi-diegetic speech“ dar, bei der ein bereits bekannter Charakter aus dem Off kommentiert.

---

[259] Vgl. Stassen (1995), S. 8
[260] Vgl. Sonnenschein (2001), S. 32 f.
[261] Vgl. Ebenda, S. 32 f.
[262] Vgl. Chion (1999), S.5
[263] Vgl. Stam / Burgoyne / Flitterman-Lewis (1994), S. 60

3. Empirie

3.1. Aktueller Forschungsstand

Das Hauptproblem der Stimmforschung stellt bis heute dar, dass es keine gesicherten Erkenntnisse über die genauen Stimmveränderungen bei emotionalen Inhalten gibt, außerdem fehlen standardisierte Tests zur Feststellung ebendieser. Zur Beobachtung emotionaler Veränderungen in Gesichtern kommt etwa die Facial Affect Scoring Technique zum Einsatz, die das Gesicht in drei Teile strukturiert, die mit Mustern verglichen werden und anhand dieser kategorisiert werden können. In Bezug auf die Stimme gibt es keinen annähernd so weit ausgebauten Test zur unabhängigen Emotionsüberprüfung, wie Tischer[264] festhält. Anhand verschiedener Untersuchungen ist zwar inzwischen zweifelsfrei geklärt, dass Emotionen über die Stimme kodiert und von Zuhörern wahrgenommen werden, doch welche Faktoren besonders auf der Senderseite dabei zum Tragen kommen und wie sich die Veränderungen einzelner Emotionen in Gehirn und Organsystem auf die Stimme im Einzelnen auswirken, ist wissenschaftlich unzureichend belegt.

Tischer[265] verglich 53 empirische Studien zum vokalen Ausdruck von Gefühlen – sowohl natürlich als auch induziert und simuliert. Dabei kamen zahlreiche Übereinstimmungen zutage, die die stimmlichen Veränderungen bei verschiedenen Emotionen belegen. Am überzeugendsten belegt (mindestens drei Studien stimmen überein, bzw. mindestens drei Studien mehr mit einem Befund als einem anderslautenden) sind zum Beispiel Veränderungen in der Lautstärke, mit der bei bestimmten Emotionen gesprochen wird. Bei Wut, Fröhlichkeit und Freude ist diese hoch, bei Unsicherheit, Traurigkeit, Langeweile und Zuneigung hingegen niedrig. Beim Sprechtempo konnten Langeweile und Traurigkeit als überwiegend langsam eingestuft werden, Interesse hingegen als schnell. Die Grundfrequenz der Stimme ist bei Wut, Angst, Fröhlichkeit, Freude und Überraschung hoch, bei Traurigkeit und Langeweile dagegen tief. Die Streubreite der Grundfrequenz ist bei Wut, Angst und Freude groß, bei Traurigkeit und Langeweile gering. Der Verlauf der Grundfrequenz wurde bei Vorwurf und Freude als unspezifisch identifiziert, bei Angst steigt, bei Traurigkeit fällt diese. Die Energie der Obertöne ist bei Ekel hoch, bei Traurigkeit jedoch gering. Die Spezifika zahlreicher weiterer Emotionen und Aspekte wie etwa Pausen- oder Vokaldauer, Lautstärkeverlauf oder Stetigkeit des Sprechens sind jedoch trotz der 53 Studien unklar oder unzureichend belegt.

---

[264] Vgl. Tischer (1993), S. 17
[265] Vgl. Ebenda, S. 83 ff.

Ein Einblick in die Rezipientenseite, wie Emotionen anhand von Stimmen dekodiert werden, ist hingegen in den letzten Jahren vermehrt gelungen, vor allem mit Hilfe der funktionellen Magnetresonanztherapie, die einen Blick in die Gehirnaktivität ermöglicht. So haben etwa Wildgruber et al.[266] im Jahr 2004bestimmte Gehirnregionen zur Verarbeitung von emotionalen Inhalten bestimmt, die bei einer rein phonetischen Prozessierung nicht zum Tragen kamen. Während bei der Unterscheidung von Vokalen eine signifikante Steigerung der Tätigkeit in der linken Gehirnhälfte (dem Frontallappen) sowie dem Scheitellappen festgestellt wurde (was auf hohe Aufmerksamkeit und Reizwechsel schließen lässt), kam bei der aktiven Verarbeitung des emotionalen Inhalts verstärkt der rechte Temporallappen – in dem der primäre auditorische Cortex und das Wernicke-Zentrum zum Sprachverständnis liegen – zum Einsatz.
Bei einer Untersuchung 2007verglichen Kreifelts et al.[267] Beobachtungen der Gehirnaktivität bei Teilnehmern, die Stimuli (sechs Emotionen sowie neutral gesprochene Aussagen) lediglich auditiv, rein visuell oder audiovisuell erhalten hatten. Sie stellen fest, dass sich die Wissenschaft lediglich auf einzelne der Felder konzentriert, die Auswirkungen des Zusammenspiels, die Unterschiede in der Wahrnehmung allerdings kaum erforscht sind. Die Audio-Gruppe konnte 56 Prozent der Emotionen korrekt identifizieren, die Video-Gruppe 75 Prozent und diejenigen Teilnehmer mit audiovisuellen Stimuli erreichten 86 Prozent. Diesen Erkenntnissen zufolge erhöht sich die Gehirnaktivität bei der audiovisuellen Stimulation im Vergleich zur einfachen deutlich, einher ging diese gesteigerte Aktivierung mit korrekt eingeordneten Emotionen. Deutlich mehr Information wurde im linken Temporallappen verarbeitet, wenn statt rein auditiver oder visueller Stimuli audiovisuelle zum Einsatz kamen. Audiovisuelle Information verstärkte die Verarbeitungsprozesse in der rechten Inselrinde, insbesondere dann, wenn sie inkongruent war. Dieser Teil des Gehirns scheint also maßgeblich daran beteiligt zu sein, Stimuli aus verschiedenen Wahrnehmungskanälen zusammenzuführen und zu vergleichen. Wie bereits bei vorherigen Forschungen (siehe 2.3.1.1) wurde auch bei diesen Versuchen eine verstärkte Aktivität in der Amygdala bei stark emotionalen, negativen Inhalten festgestellt, etwa beim Hören einer ängstlichen Stimme.

Die Auswirkungen von Emotionen verschiedener Valenz und Intensität erforschten Ethofer et al.[268] im Jahr 2005, da auch sie der Ansicht waren, dass sich lediglich wenige Studien mit der

[266] Vgl. Wildgruber et al., in: NeuroImage (2005), S. 1233 ff.
[267] Vgl. Kreifelts in: NeuroImage (2007), S. 1445 ff.
[268] Vgl. Ethofer et al., in: NeuroReport (2006), S. 249 ff.

Verarbeitung emotionaler Informationen über die Stimme beschäftigen. Bei der Untersuchung wurde festgestellt, dass das Wernicke-Sprachzentrum sowohl auf stark fröhliche wie auch stark wütende Stimmeindrücke verstärkt reagiert. Da lediglich diese beiden Emotionen getestet wurden, konnte jedoch keine generalisierende Aussage über die Erlebnisqualität negativer und positiver Emotionen getroffen werden. Die Autoren weisen außerdem darauf hin, dass sich im emotionalen Ausdruck von Emotionen von einer Stimme zur anderen enorme Unterschiede abzeichnen. Um die Auswirkungen dieser Besonderheiten zu erforschen, wurden Aussagen technisch verändert und Durchschnittsstimmen gebildet, ebenso wie Veränderungen der Lautstärke und der Frequenz vorgenommen. Wie sich herausstellte, hatte keine dieser Veränderungen Auswirkungen auf die emotionale Verarbeitung im Gehirn – emotionale Eindrücke können demnach generalisiert werden und sind nicht oder nur zum Teil von interpersonellen Unterschieden in der Stimme abhängig. Bei wütenden Stimmen zeigte sich beidseitig eine verstärkte Aktivität des mittleren Temporallappens – wo sich unter anderem der Hippocampus (ein Teil des Sitzes des Gedächtnisses) befindet. Künstlich veränderte Klänge, die den Spezifika wütender Stimmgebung entsprachen, führten jedoch zu keiner verstärkten Aktivierung dieses Teils des Gehirns. Insgesamt reagierte der rechte Teil des mittleren Temporrallappens stärker auf den emotionalen Ausdruck als der linke.

Wie die Emotionen verpackt sind, scheint Einfluss auf die Wahrnehmungsgenauigkeit zu haben, wie Tischer[269] in einer Studie herausgefunden hat. Positive Gefühle werden besser erkannt, wenn sie in Sätze gekleidet sind, die von finalem Intonationsabfall und Vokaldehnung der Silben gekennzeichnet sind. Negative Gefühle hingegen können auch ohne finalen Intonationsabfall erfasst werden. Je länger die Aussage dauert, die zur Identifikation des Sprechergefühls dient, desto sicherer können Zuhörer die Valenzdimension des Gehörten bestimmen.

3.2. Forschungsfragen und Hypothesen

Forschungsfrage 1: Können Zuhörer anhand eines Serien-Ausschnitts, der aus einer stimmlichen Äußerung von nur wenigen Sekunden besteht, erkennen, welche Emotion/en vom Synchronsprecher vermittelt wird/werden?
Hypothese 1.0: Zuhörer können anhand eines kurzen Ausschnitts nicht erkennen, welche Emotion/en vermittelt wird/werden.

[269] Vgl. Tischer (1993), S. 318 ff.

Hypothese 1.1: Zuhörer können anhand eines kurzen Ausschnitts erkennen, welche Emotion/en vermittelt wird/werden.

Forschungsfrage 2: Welche Emotionen werden von Zuhörern besonders klar erkannt und welche sind schwer zu identifizieren?
Hypothese 2.0: Emotionen werden unabhängig von ihrer Valenz in ähnlichem Ausmaß erkannt.
Hypothese 2.1: Stark negative Emotionen werden im Allgemeinen besser erkannt als positive und neutrale Emotionen.
Hypothese 2.2: Positive Emotionen werden im Allgemeinen besser erkannt als neutrale Emotionen.

Forschungsfrage 3: Wodurch unterscheiden sich die Ergebnisse zwischen Zuhörern und Zusehern, denen nur das zur Stimme gehörige Bild ohne Ton vorgelegt wird?
Hypothese 3.0: Emotionen werden im Allgemeinen gleich gut erkannt**.**
Hypothese 3.1: Zuhörer erkennen Emotionen besser als die Zuseher.
Hypothese 3.2: Zuhörer können stark negative Emotionen besser erkennen als Zuseher.
Hypothese 3.3: Zuhörer können stark positive Emotionen besser erkennen als Zuseher.

Forschungsfrage 4: Welche Unterschiede zeigen sich in den Ergebnissen zwischen reinen Zuhörern, reinen Zusehern sowie jenen Personen, die die betreffenden Ausschnitte mit Bild und Ton sehen?
Hypothese 4.0: Emotionen werden im Allgemeinen gleich gut erkannt.
Hypothese 4.1: Wer den Ausschnitt audiovisuell erlebt, kann Emotion/en besser zuordnen als reine Zuhörer und Zuseher.
Hypothese 4.2: Wer den Ausschnitt audiovisuell erlebt, kann stark negative Emotionen besser unterscheiden als reine Zuhörer und Zuseher.
Hypothese 4.3: Wer den Ausschnitt audiovisuell erlebt, kann stark positive Emotionen besser unterscheiden als reine Zuhörer und Zuseher.

Forschungsfrage 5: Welche Unterschiede lassen sich in der auditiven Emotionswahrnehmung in Verbindung mit Persönlichkeitsmerkmalen sowie Empathiefähigkeit ermitteln?

Hypothese 5.0: Es gibt keinen Zusammenhang zwischen Persönlichkeitsmerkmalen und der Erkennung von Emotionen.
Hypothese 5.1: Personen mit niedrigen Werten im Bereich Extraversion können Emotionen besser zuordnen als solche mit hohen.
Hypothese 5.2: Personen mit hohen Werten im Bereich Emotionalität können Emotionen besser zuordnen als solche mit niedrigen.
Hypothese 5.3: Personen mit hohen Werten im Bereich Lebenszufriedenheit können Emotionen besser zuordnen als solche mit niedrigen.
Hypothese 5.4: Personen mit hohen Werten im Bereich Soziale Orientierung können Emotionen besser zuordnen als solche mit niedrigen.
Hypothese 5.5: Personen mit hohen Werten im Bereich Gehemmtheit können Emotionen schlechter zuordnen als solche mit niedrigen.
Hypothese 5.6: Personen mit hohen Werten im Bereich Empathie in realen Situationen können Emotionen besser zuordnen als solche mit niedrigen.
Hypothese 5.7: Personen mit hohen Werten im Bereich Empathie in fiktiven Situationen können Emotionen besser zuordnen als solche mit niedrigen.

Forschungsfrage 6: Welche Unterschiede lassen sich in der auditiven Emotionswahrnehmung in Verbindung mit sozialen Daten ermitteln?
Hypothese 6.0: Es gibt keinen Zusammenhang zwischen sozialen Daten und der Emotionserkennung.
Hypothese 6.1: Frauen können Emotionen besser zuordnen als Männer.
Hypothese 6.2: Ältere Menschen können Emotionen besser zuordnen als jüngere.
Hypothese 6.3: Personen mit Geschwistern können Emotionen besser zuordnen als Einzelkinder.
Hypothese 6.4: Personen mit Kindern können Emotionen besser zuordnen als kinderlose Menschen.
Hypothese 6.5: Eltern von Töchtern können Emotionen besser zuordnen als Eltern von Söhnen.
Hypothese 6.6: Personen, die in einer Beziehung leben, können Emotionen besser zuordnen als Singles.

Forschungsfrage 7: Welchen Einfluss hat der Faktor, ob eine Person eine TV-Serie bereits mindestens einmal gesehen hat, auf die Erkennung der von der Synchronstimme dargestellten Emotion?

Hypothese 7.0: Es besteht kein Unterschied in Bezug auf die Genauigkeit der Emotionserkennung zwischen Menschen, die eine TV-Serie bereits gesehen haben und jenen, die diese Serie noch nicht gesehen haben.

Hypothese 7.1: Zuhörer, die eine TV-Serie bereits gesehen haben, können die Emotion besser zuordnen als Personen, die die Serie noch nie gesehen haben.

Hypothese 7.2: Je öfter eine Person die TV-Serie gesehen hat, desto besser kann sie die Emotion zuordnen.

3.3. Methodisches Design

Bei dieser Untersuchung handelte es sich um eine quantitative Querschnittserhebung mit einfacher Zufallsstichprobe, die allerdings anhand der Schneeballtechnik gewonnen wurde. Die Untersuchung bestand aus verschiedenen Fragebögen, darunter Teile des Freiburger Persönlichkeitsinventars sowie der Empathie-Skala und einem selbst entworfenen multimedialen Fragebogen, dessen Bestandteile anhand der qualitativen Inhaltsanalyse nach Hickethier ausgewählt wurden. Der vollständige Fragebogen sowie das Einstellungsprotokoll der ausgewählten Stimuli finden sich im Anhang.

3.3.1. Einstellungsanalyse nach Hickethier

Hickethier[270] empfiehlt zur Protokollierung von Filmen oder Teilen davon eine qualitative Inhaltsanalyse, bestehend aus Sequenzliste und/oder Einstellungsprotokoll. In der Sequenzliste werden Handlungseinheiten beschrieben, die meist aus mehreren Einstellungen bestehen. Der Beginn einer neuen Sequenz wird dabei meist durch einen Wechsel des Ortes oder der erzählten bzw. Erzählzeit markiert. Die Einstellungsanalyse ist das feine Messinstrument für eine einzelne Sequenz, hierbei wird jede Einstellung beschrieben, sowie Dialoge und Off-Texte. Das Protokoll teilt sich in verschiedene Spalten für Bild und Ton sowie die Dauer der jeweiligen Einstellung, dies soll gestalterische Details deutlich machen, es ermöglicht eine gründliche Auseinandersetzung mit dem Film und lässt Einzelheiten sowie stilistische Merkmale, die ansonsten bei normaler Betrachtung kaum Beachtung finden, erkennen.

[270] Vgl. Hickethier (1996), S. 38 f.

- **Vorteile**

Die Vorteile des Einstellungsprotokolls nach Hickethier liegen in dieser Arbeit vor allem darin, dass es eine gleichbleibende Größe bei der Analyse eines filmischen Inhalts ist – es ermöglichen einen unabhängigen Blick auf verschiedenartige Inhalte und macht sie so vergleichbar. So fiel es leichter, geeignete Szenen auszuwählen, die sowohl für reine Zuhörer oder Zuseher als auch für audiovisuelle Zuseher geeignet waren um Emotionen zu erkennen.

- **Nachteile**

Ein möglicher Nachteil an dieser Methode ist der Zeitaufwand, der mit der sekundengenauen Erfassung der beiden Protokolle sowie der genauen Auflistung der einzelnen Einstellungen oder Sequenzen verbunden ist. Für eine wissenschaftliche Untersuchung ist dieser Aufwand jedoch unabdingbar.

- **Bedingungen**

Grundvoraussetzung für die Film- (bzw. in diesem Fall Fernseh-) Analyse nach Hickethier ist ein vollständiges, fehlerloses Vorliegen des Materials. Um die Protokolle sinnvoll (also sekundengenau) erstellen zu können, ist besonders exakte Audiovideo-Software vonnöten, um auch Überblendungen sowie Schnitte in der Kürze von Millisekunden zu erkennen und auch jede Sekunde des filmischen Ausschnitts zerlegen und genau betrachten zu können. Bei dieser Arbeit kam dabei das Freeware-Programm (legal und gratis über das Internet herunterzuladen) „Avidemux“ zum Einsatz. Es bietet diverse Möglichkeiten, die eine exakte Analyse ermöglichen – zum Beispiel können auf Wunsch Millisekunden genau Sequenzen aus dem filmischen Material extrahiert werden. Außerdem ist ein Sprung innerhalb des Materials von nur wenigen Millisekunden möglich, was vor allem bei der visuellen Begutachtung hilft – so werden sämtliche Einstellungen, Überblendungen und Schnitte sichtbar.

3.3.2. Fragebögen

3.3.2.1. Empathie-Skala

Die Empathie-Skala, kurz E-SKala, nach Leibetseder et al.[271] wurde aus verschiedenen Items englischsprachiger Tests zur Empathiemessung zusammengesetzt. Empathie wird hier als Befähigung definiert, Gefühle anderer erkennen und nachempfinden zu können. Die E-SKala besteht aus 25 Items, die auf einer Skala von „Trifft zu“ bis „Trifft nicht zu“ bewertet werden, hohe Werte entsprechen bis auf zwei negativ gepolte Items hoher Empathiefähigkeit. Der

[271] Vgl. Leibetseder et al., in: Zeitschrift für Differentielle und Diagnostische Psychologie (2001), S. 70 ff.

Fragebogen setzt sich aus zwei Teilen zusammen, ersterer dreht sich um die Phantasieempathie, also die Einfühlungsbereitschaft in fiktive Situationen – Bücher, Filme und Fernsehen. Teil zwei behandelt die Betroffenheit in direkter interpersonaler Kommunikation, die Empathie in realen Situationen.

- **Vorteile**

Diese Empathie-Skala eignet sich besonders gut, wenn ein Vergleich zwischen der Empathiefähigkeit in fiktiven und realen Situationen interessant ist, da keine unterschiedlichen Fragebögen verwendet werden müssen und der Flow beim Ausfüllen so nicht unterbrochen wird. Beim Test der E-SKala haben sich signifikante Korrelationen mit verschiedenen anderen Fragebögen zur Erhebung von Empathie, Einfühlungsbereitschaft und Anteilnahme gezeigt, was auf eine hohe Validität und Stabilität schließen lässt.

- **Nachteile**

Die Empathie-Skala kann nicht mit einer repräsentativen Umfrage aufwarten, getestet wurde sie anhand einer Stichprobe von 141 Personen, mit Testwiederholung bei 53 Personen. Auch zählt sie noch nicht zum wissenschaftlichen Standard, die Ergebnisse sind daher nur beschränkt interpretier- und vergleichbar.

- **Bedingungen**

Die Intensitätsskala bei den Antwortmöglichkeiten erleichtert die Auswertung der beiden Faktoren Empathie in fiktiven und realen Situationen sowie deren Vergleichbarkeit. Um die Unterscheidung deutlicher zu gestalten, wurde für die vorliegende Arbeit auf Anraten des Betreuungsteams von Prof. Dr. Vitouch die Skala von „Trifft zu“ bis „Trifft nicht zu“ auf „Trifft sehr zu“ bis „Trifft gar nicht zu“ geändert.

3.3.2.2. Das Freiburger Persönlichkeitsinventar

In meiner Untersuchung kamen Teile des Freiburger Persönlichkeitsinventars (FPI-R) zum Einsatz, ich habe 56 von insgesamt 137 Items verwendet, diese entsprechen fünf von zwölf Skalen des FPI-R.

Das Freiburger Persönlichkeitsinventar ist ein wissenschaftlicher Test, der verschiedene Persönlichkeitsmerkmale anhand von137 Items misst. Er dient sowohl zum Vergleich großer Personengruppen als auch als individueller psychologischer Test. Das Freiburger

Persönlichkeitsinventar wurde in den 1960er Jahren entwickelt und für die 7. Auflage 2001 nach einer repräsentativen Umfrage in Deutschland mit 3.740 Teilnehmern neu normiert. 138 Items (das erste soll jedoch lediglich die Motivation erhöhen und sozial erwünschte Aussagen vermindern, wird jedoch nicht ausgewertet) werden in zehn Standard- und zwei Zusatzskalen eingeteilt. Die Standardskalen sind: Lebenszufriedenheit, Soziale Orientierung, Gehemmtheit (diese drei wurden von mir eingesetzt) sowie Leistungsorientierung, Erregbarkeit, Aggressivität, Beanspruchung, Körperliche Beschwerden, Gesundheitssorgen und Offenheit. Die zwei Zusatzskalen, die von mir ebenfalls verwendet wurden, beschreiben Extraversion und Emotionalität.[272]

Ausgewertet wird das FPI-R anhand eines Itemschlüssels, der hauptsächlich „stimmt", aber auch einige „stimmt nicht"-Antworten zählt und den verschiedenen Skalen zuordnet. Diese Ergebnisse sind Rohwerte, die mit Hilfe von Statinen (Standard Nine) normiert werden. Diese Normen finden sich in der Handanweisung des FPI-R, sie sind in Geschlecht sowie Altersgruppen eingeteilt. Der Mittelwert dieser Normen liegt bei 5, die Standardabweichung bei 1,96. In 54 % der Fälle treten laut FPI-R Werte von 4 bis 6 auf der Statine-Skala (von 1 bis 9) auf.[273]

- **Vorteile**

Persönlichkeitsinventare im Allgemeinen bieten eine gute Möglichkeit, eine große Anzahl Befragter in Bezug auf ihre Persönlichkeit zu vergleichen. Mit relativ wenig Aufwand lassen sich viele Persönlichkeitsmerkmale einer Person abfragen. Durch die Einteilung der Items in Skalen können den Teilnehmern Ränge auf ebendiesen Skalen zugeordnet werden. Persönlichkeitsinventare lassen nicht nur einen Schluss auf Persönlichkeitsmerkmale sondern auch auf konkretes Verhalten zu.[274]

Das Freiburger Persönlichkeitsinventar kam in meiner Untersuchung auch deshalb zum Einsatz, weil es mehrfach anhand repräsentativer Umfragen überarbeitet und streng getestet wurde, mit zahlreichen anderen Persönlichkeitstests korreliert, die Einteilung der Skalen dem Forschungsinteresse entgegen kommt und sich das dichotome Antwortformat („Stimmt" und „Stimmt nicht") in einem Pretest als vorteilhaft herausstellte.

- **Nachteile**

[272] Vgl. Fahrenberg / Hampel / Selg (2001), S. 7 ff.
[273] Vgl. Ebenda, S. 80 ff.
[274] Vgl. Fahrenberg / Hampel / Selg (2001), S. 8 ff.

Kritik am FPI-R (sowie an zahlreichen anderen Persönlichkeitsinventaren) gibt es an der geringeren Stabilität ($.50 < r_{tt} < .70$) gegenüber Leistungstests, der geringeren inneren Konsistenz ($.60 < r_{tt} < .80$) sowie teils geringer Korrelation ($r = .20$ bis $r = .30$) mit quantitativ abgestuften Außenkriterien. Eine unerwünschte Möglichkeit stellen sozial erwünschte Antworten bzw. die „Lügenskala“ dar. Diese Punkte werden jedoch bei zahlreichen psychologischen Tests als problematisch angesehen oder haben ihrerseits Kritiker auf den Plan gerufen, die etwa auf die schwierige Vergleichbarkeit mit anderen Tests hinweisen oder Korrelationskoeffizienten als alleiniges Kriterium nicht gelten lassen.[275]

- **Bedingungen**

Zur Verwendung einzelner Skalen des Freiburger Persönlichkeitsinventars erklären die Autoren[276], dass das FPI-R in der Gesamtheit normiert und standardisiert wurde, eine Trennung daher möglicherweise problematisch sein und Einfluss auf die Ergebnisse haben könnte. Gleichzeitig stellen sie aber fest, dass die Position der Items und Frageblöcke laut verschiedener Untersuchungen erheblich weniger ausschlaggebend ist als postuliert. Vor Beginn muss eine Testinstruktion ausgegeben werden, die vor allem darauf hinweist, dass es keine richtigen oder falschen Antworten gibt und über die Fragen nicht lange nachgedacht werden sollte. Für die Auswertung gilt selbstverständlich, dass diese professionell und nach den Angaben des FPI-R erfolgen muss, die Daten streng vertraulich zu behandeln sind und von nicht befugten Personen keine Persönlichkeitsprofile einzelner Personen erstellt werden dürfen.

- **Verwendete Skalen**

Lebenszufriedenheit: Ein hoher Testwert entspricht hoher Zufriedenheit mit dem eigenen Leben, der Partnerschaft und Zuversicht, ein niedriger zeugt von Unzufriedenheit mit den eigenen Umständen, der Beziehung sowie Depressivität.
Soziale Orientierung: Personen mit hohen Werten fühlen sich sozial verantwortlich, sind hilfsbereit und möchten ihr Wohlbefinden sowie ihren Wohlstand mit anderen teilen. Niedrige Werte sprechen von Selbstbezogenheit, wenig Solidaritäts- und Mitgefühl sowie dem Unverständnis gegenüber sozialen Einrichtungen.
Gehemmtheit: Hohe Werte bedeuten bei dieser Skala Unsicherheit im Umgang mit anderen Menschen, Gehemmtheit und die Vermeidung sozialer Kontakte. Niedrige Werte zeugen von einer selbstsicheren, ungezwungenen Persönlichkeit, die gerne auf andere zugeht.

[275] Vgl. Ebenda, S. 9 ff.
[276] Vgl. Ebenda, S. 80

Extraversion: Hohe Testwerte stehen für extrovertierte, unternehmungslustige Menschen, die gerne in geselliger Runde unterwegs sind und spontane Entscheidungen treffen. Im Gegensatz dazu bedeuten niedrige Werte, dass die Person introvertiert und ernst ist, sich zurückhält und sehr überlegt handelt.
Emotionalität: Personen mit hohen Werten fühlen sich von zahlreichen Problemen geplagt, oft gepaart mit körperlichen Beschwerden, sie sind ängstlich und labil. Niedrige Werte stehen dagegen für emotionale Stabilität, Selbstvertrauen, Zufriedenheit und einen gelassenen Umgang mit dem Leben.[277]

3.3.2.3. Multimedialer Fragebogen

Den Kern meiner Untersuchung bildete ein multimedialer Fragebogen. Hier waren verschiedene reine Audio- oder Video- sowie audiovisuelle Dateien eingebunden, darunter befand sich die zugehörige Frage mit den Antwortmöglichkeiten.

- **Vorteile**

Für das Ausfüllen eines multimedialen Fragebogens ist lediglich ein Computer mit angeschaltetem Sound nötig. Ein multimedialer Fragebogen individualisiert das Sehen von auditiven, visuellen und audiovisuellen Dateien, im Gegensatz etwa zu Gruppenvorführungen auf einer Leinwand, wo durch die Reaktionen anderer Zuseher das Untersuchungsergebnis möglicherweise verfälscht werden kann. Ein weiterer Vorteil liegt in der Auswertbarkeit der Ergebnisse, da bereits im Vorfeld Missing Values definiert werden und die gespeicherten Daten relativ einfach in ein Datenauswertungsprogramm (in diesem Fall SPSS) importiert werden können.

- **Nachteile**

Die Erstellung eines multimedialen Fragebogens erfordert Programmierkenntnisse, die zum Beispiel bei der traditionellen Form eines Fragebogens auf Papier und dem Vorspielen der audiovisuellen Dateien auf einem Fernseher oder einer Leinwand nicht vonnöten sind. Die Teilnehmer dürfen auch keine Scheu vor dem Bedienen eines Computers besitzen, müssen zumindest über rudimentäre Kenntnisse in diesem Bereich verfügen.

- **Bedingungen**

[277] Vgl. Fahrenberg / Hampel / Selg (2001), S. 72 ff.

Neben dem Know-How sind für die Erstellung eines multimedialen Fragebogens diverse Computerprogramme nötig, etwa zum Schneiden von Audio- und Videosequenzen (in diesem Fall Avidemux) oder zum Erstellen des Fragebogens an sich (hier kam vor allem das Freeware-Programm Weaverslave, das gratis im Internet zum Download bereitsteht, zum Einsatz).

3.3.3. Untersuchung via Internet

Meine Untersuchung wurde im Internet durchgeführt, wo ich die verschiedenen Fragebögen zu einer Befragung zusammengeführt habe.

- **Vorteile**

Die Vorzüge einer Befragung via Internet liegen in der einfachen, schnellen Verfügbarkeit. Die Versuchspersonen sind nicht ortsgebunden und unterliegen keinem Zeitdruck, was die Motivation steigern und eine größere Anzahl an Befragungsteilnehmern einbringen sollte. Auch die Einbindung multimedialer Dateien in die Befragung ist dank „YouTube“ oder ähnlichen Online-Diensten ohne Unterbrechung möglich, so wird das Flow-Erlebnis beim Ausfüllen nicht gestört – im Gegensatz dazu, wenn z.B. ein Filmausschnitt auf einer Leinwand gezeigt wird, anschließend Fragebögen ausgeteilt werden und so eine zeitlicher Einschnitt zwischen der Rezeption und der Bewertung des Gesehenen entsteht.

- **Nachteile**

Der vielleicht größte Nachteil eines Online-Fragebogens ist die technische Umsetzung, die nicht unerhebliche Programmierkenntnisse verlangt – im konkreten Fall vor allem die automatische, zufällige Einteilung in drei verschiedene Untersuchungsgruppen. Schließlich wurde das Ziel zwar erreicht, allerdings konnten nur Personen mit einem „Firefox“-Internetbrowser an der Befragung teilnehmen, da die Programmierung den Dienst beim „Internet Explorer“ von Microsoft den Dienst verweigerte. Damit ist – trotz eindringlicher Erklärung vor Start des Fragebogens – die hohe Rücklaufquote an nicht oder nicht vollständig ausgefüllten Fragebögen zu erklären: Viele Personen (vor allem ältere) wussten mit dem Begriff „Firefox“ trotz Erklärung und Link zum Download nichts anzufangen und besaßen nicht genug technische Kenntnisse, um den alternativen Browser zu installieren.
Ein weiterer Nachteil des Online-Fragebogens ist die geringere Hemmung, den Fragebogen halb ausgefüllt abzubrechen, da der Versuchsleiter nicht motivierend oder kontrollierend auf

den Befragungsteilnehmer einwirken kann. Auch das ist eine mögliche Erklärung für die hohe Rücklaufquote an nicht verwertbaren Fragebögen.

- **Bedingungen**

Um eine Befragung im Internet durchführen zu können, ist ein Server (in diesem Fall des Studentenaccounts der Universität Wien) für die zugehörigen Dateien vonnöten, außerdem eine derart programmierte Website, dass die Befragten flüssig von einem zum nächsten Teil der Untersuchung weitergeleitet und ihre Antworten ebenfalls auf einem Server gespeichert werden.

3.4. Praktische Umsetzung

Um herauszufinden, ob Emotionen nur anhand von stimmlichen Äußerungen von wenigen Sekunden Länge und ohne visuelle Unterstützung erkennbar sind, wurden den Versuchspersonen bei dieser Untersuchung Ausschnitte von auf Deutsch synchronisierten Fernsehserien vorgelegt. Jedem der sieben Ausschnitte mussten die Befragten anschließend eine oder zwei Emotionen zuordnen, die sie als zugehörig empfanden. Es standen jeweils acht Emotionen zur Auswahl, die Basisemotionen nach Plutchik. Diese Emotionseinteilung wurde ausgewählt, da acht Emotionen als nicht zu beschränkte aber auch nicht zu zahlreiche Auswahl erschien. Die Befragten sollten mit dieser Anzahl weder unter- noch überfordert sein, dennoch sollte sie differenzierte Ergebnisse ermöglichen. Plutchik[278] selbst weist die Verwendung von Emotionswörtern zur Beschreibung von Emotionen in seiner Theorie hin, was die Verwendung seines Ansatzes ebenfalls sinnvoll machte.
Um die Ergebnisse in einen Zusammenhang stellen zu können, wurde jeder Ausschnitt je einer Gruppe lediglich als Audio-Datei vorgelegt, der ersten Kontrollgruppe als rein visuelle Datei (ohne Ton) und der zweiten Kontrollgruppe als audiovisuelle Datei (also so, wie es der üblichen TV-Konsumation entspricht). Jede Gruppe erhielt sowohl reine Audio- als auch Video- sowie Audiovideo-Dateien.
Zu Beginn der praktischen Untersuchung stand die Auswahl der Ausschnitte, die die Testpersonen zu sehen und zu hören bekommen sollten. Zu diesem Zweck wurden mehrere Folgen verschiedener synchronisierter US-Fernsehserien aufgezeichnet und anschließend kritisch in Bezug auf mehrere Kriterien untersucht. Der Entschluss, synchronisierte Fernsehserien zu verwenden, fiel aus mehreren Gründen:

[278] Vgl. Plutchik (2003), S. 93 ff.

- Fernsehen gehört für den Großteil der westlichen Welt zum täglichen Leben. Es kann daher davon ausgegangen werden, dass der Großteil der Befragungsteilnehmer mit dem Medium Fernsehen vertraut sein würde und die Untersuchung deshalb Rückschlüsse auf Emotionserkennung im Medium Fernsehen als Ganzes zulässt.
- Synchronisierte Fernsehserien gehören in deutschsprachigen Ländern ebenfalls zum Standard – meist handelt es sich dabei um Serien aus den USA. Daher ist auch davon auszugehen, dass der überwiegende Teil der Bevölkerung mit Synchronisation (also zum Beispiel dem Fakt, dass die Stimme nicht immer lippensynchron erklingt) vertraut ist – wenn auch unterbewusst.
- Große US-TV-Produktionen werden hochwertig und kostspielig produziert. Daher wird auch großer Wert auf gute Synchronsprecher gelegt, die das Gefühl der Serie im Original vermitteln und der Rolle auch ohne die eigentliche Stimme des Schauspielers Bedeutung verleihen. Synchronsprecher unterliegen also einer strengen Auswahl – was den Rückschluss erlaubt, dass ihre Stimmen und Sprechweisen eine besondere Güte besitzen, dass sie Stimmungen und Emotionen gut transportieren können. Zudem kann sich der Synchronsprecher auf seine stimmliche Arbeit konzentrieren, er muss sich „lediglich“ an die Vorgaben des Tonmeisters und des Schnitts halten – hat jedoch nicht wie ein Schauspieler noch unzählige weitere Aufgaben wie auf seine Körperhaltung, das Licht, die Kamera und andere Akteure zu achten. Diese Konzentration auf die stimmliche Arbeit des Synchronsprechers sollte helfen, auch aus sehr kurzen Ausschnitten klare Emotionen herausfiltern zu können.

Dass letztendlich mit „CSI: Miami“ und „Damages – Im Netz der Macht“ zwei – wenn auch sehr unterschiedliche – Krimi-Serien ausgewählt wurden, ergab sich im Laufe des Sichtungsprozesses zahlreicher verschiedener TV-Serien.

„CSI: Miami“ läuft seit 2004 (mit Unterbrechungen) im österreichischen Fernsehen – sowohl im öffentlich-rechtlichen als auch auf privaten Sendern. Die Serie zählt zu den erfolgreichsten und bekanntesten Krimi-Serien der Welt, 2006 wurde „CSI: Miami“ in einer Studie von Informa Telecoms and Media, wie „BBC Online“[279] berichtete, in 20 Ländern und damit weltweit zur beliebtesten Fernsehserie gekürt. Von „Damages“ war bisher (im Jahr 2008) nur die erste Staffel im ORF zu sehen – zu nachtschlafender Zeit. Dieser Unterschied in der Verfügbarkeit sollte helfen herauszufinden, ob die Emotionserkennung bei bekannten Serien

[279] Vgl. BBC Online (2006), URL: http://news.bbc.co.uk/1/hi/entertainment/5231334.stm

leichter fällt als bei weniger bekannten – ob die Vertrautheit mit Schauspielern und deren Synchronsprechern Einfluss auf die Emotionserkennung hat.
Im Vordergrund bei der Auswahl der Serien und der letztendlich gezeigten Ausschnitte stand die Sprachverständlichkeit sowie dass das Gesicht der sprechenden Person im Bild zu sehen sein musste. Schließlich wurden neben der rein auditiven Datei auch eine rein visuelle sowie eine audiovisuelle Kopie des Ausschnitts erzeugt.

3.4.1. Inhaltsanalyse

Nach intensiver Auseinandersetzung mit verschiedenen Fernsehserien (in der Vorauswahl standen unter anderem auch „Desperate Housewives“, „Men in Trees“, „The Closer“ und „Las Vegas“) fiel die Entscheidung für „CSI: Miami“ und „Damages“. Beide Serien setzen wenig Hintergrundmusik ein, außerdem ist die Umgebungslautstärke meist sehr gering – aus diesem Grund fiel „Las Vegas“ aus der Liste. Der visuelle Code beider Serien legt außerdem Wert auf die Darstellung des Gesichts jener Person, die gerade spricht – was für diese Untersuchung unabdingbar ist. „The Closer“ hätte zwar sehr emotionale Szenen geboten, allerdings ist dort meist die angesprochene Person und ihre Reaktion auf das Gesagte zu sehen statt der Sprecher selbst – zudem wird meist inmitten von Sätzen die Perspektive gewechselt. Außerdem wurde im Lauf des Sichtungsprozesses deutlich, dass lustige Serien wie „Desperate Housewives“ und „Men in Trees“ keine geeigneten Szenen mit starken Emotionen in der Stimme liefern – da in diesen Serien Emotionen meist stark übertrieben dargestellt werden.
Mehrere Szenen aus „CSI: Miami“ und „Damages – Im Netz der Macht“ sind jedoch allen Ansprüchen gerecht geworden, daher wurden verschiedene aufgezeichneten Folgen dieser Serien einer Inhaltsanalyse unterzogen. Diese Inhaltsanalysen nach Hickethier finden Sie im Anhang.

Ausgewählt wurden schließlich sieben verschiedene kurze Ausschnitte, die aus Szenen stammen, die eindeutig eine oder zwei Emotion/en vermitteln – dies wurde auch in einem Pretest erhoben. Aus dem Kontext der Szene sowie der ganzen Folge wurden jene Emotionen herausgefiltert, die es für die Befragungsteilnehmer herauszuhören/sehen galt.

Die sieben kurzen Ausschnitte stammen aus diesen Folgen:
„Damages“ – 1. Staffel – Folge 5 – „Meineid“ (Original: „A regular Earl Anthony“): „Ich bin so froh, dass du da bist!“

„Damages“ – 1. Staffel – Folge 6 – „Wahre Gesichter“ (Original: „She spat at me“): „Nein, du hast gegrunzt wie ein Höhlenmensch!“ und „Warum sollten wir das nicht?“
„Damages“ – 1. Staffel – Folge 8 – „Der Verräter” (Original: „Blame the Victim“): „David!“
„CSI: Miami“ – 2. Staffel – Folge 1 – „Blutsbrüder“ (Original: „Blood Brothers“): „Ich ruf’ da jetzt nicht an!“ und „Ja, und wahrscheinlich ist es das, was mir ein Überschreiten der Grenze unmöglich macht“ und „Ich würde zu gern wissen, wie die aus deiner Wohnung hierher gekommen ist!“

Aus den Analysen der ganzen Folge sowie der Szenen, aus denen die kurzen Ausschnitte stammen, ergibt sich folgende Emotionszuweisung (zum Einsatz kam jeweils lediglich der fett markierte Satz):

- **„Damages“:**

„Ich bin so froh, dass du da bist“ (3 Sekunden) – Hauptfigur Ellen Parsons (Rose Byrne) hat Besuch von ihrer Mutter (Debra Monk), die ihr lang und breit von ihrem friedlichen Städtchen erzählt und darüber nicht merkt, wie sehr Ellen unter einem schwelenden Konflikt mit ihrem Verlobten sowie der Überforderung durch ihren neuen Job als Anwältin in einer großen Anwaltskanzlei leidet. Ellen sitzt nur noch still auf dem Bett mit dem Brautkleid auf ihrem Schoß und bricht plötzlich in Tränen aus, woraufhin ihre zuvor ahnungslose Mutter sich zu ihr setzt und sie in den Arm nimmt, was Ellen dankbar annimmt. Die Emotionen, die von Rose Byrne und ihrer Synchronsprecherin hier dargestellt werden, sind **Vertrauen** und **Traurigkeit**.
„Nein, du hast gegrunzt wie ein Höhlenmensch!“ (2 Sekunden) – Ellen war mit ihrem Verlobten David Connor (Noah Bean) bei ihrer neuen Chefin Patty und deren Ehemann zum Abendessen eingeladen. David hat sich während des ganzen Abends feindselig gegenüber den Gastgebern verhalten. Bei der Rückkehr in die eigene Wohnung ist der Streit über Davids Verhalten in vollem Gang, Davids Wortkargheit und schlechtes Benehmen sind Ellen ein Dorn im Auge. David erklärt lapidar: „Ich habe ihre (Pattys) Fragen beantwortet!“, woraufhin Ellen ihre **Wut** und den **Ekel** gegenüber seinem Verhalten zum Ausdruck bringt.
„Warum sollten wir das nicht?“ (2 Sekunden) – Ellen wird von ihrer Chefin Patty Hewes (Glenn Close) nach dem verpatzten Abendessen auf die Beziehung zu David angesprochen. Patty kommentiert die Konflikte zwischen den beiden mit den Worten „Ich hoffe, Sie beide schaffen das“, was bei Ellen **Überraschung** und **Neugierde** als Reaktionen weckt.
„David!“ (2 Sekunden) – In Rückblenden wird in jeder Folge „Damages“ der ersten Staffel Stück für Stück aufgeklärt, wie es dazu kam, dass Ellen des Mordes an ihrem Verlobten

angeklagt wird. Sie ist jedoch selbst beinahe einem Mordanschlag zum Opfer gefallen und nur knapp mit dem Leben davongekommen. In dieser Szene zeigt die Rückblende den Zeitpunkt, in dem Ellen nach dem Angriff auf sie mit Blut an Körper und Kleidung ihre Wohnung betritt. Voller **Angst**, um nicht zu sagen Panik, schreit sie nach ihrem Verlobten, den sie erschlagen in der Badewanne findet.

- **„CSI: Miami“**

„Ich ruf’ da jetzt nicht an!“ (2 Sekunden) – Eric Delko (Adam Rodriguez) wird von seinem Kollegen Tim Speedle (Rory Cochrane) unter Druck gesetzt, eine seiner Ex-Freundinnen anzurufen, um die Zulassungsdaten eines Autos zu erfahren, mit dem eine Frau überfahren wurde. Eric ist überrascht, dass sich Tim noch an die Frau erinnern kann und sträubt sich aufgrund unangenehmer Erinnerungen an die Ex-Freundin. Die dargestellten Emotionen sind **Überraschung** und **Ekel**.
„Ich würde zu gern wissen, wie die aus deiner Wohnung hierher gekommen ist!“ (4 Sekunden) – Eric Delko und Tim Speedle wurden auf ein Müllschiff abkommandiert, wo sie nach dem Airbag des fraglichen Wagens suchen. Eric ist überrascht, was sich alles zwischen dem Müll findet – zum Beispiel eine tote Ratte, die er Tim mit einem hämischen Grinsen unter die Nase hält. Die dargestellten Emotionen sind **Überraschung** und **Freude**.
„Ja, und wahrscheinlich ist es das, was mir ein Überschreiten der Grenze unmöglich macht“ (5 Sekunden)– Der Leiter der Spurensicherung, Horatio Caine (David Caruso), trifft in einem Restaurant auf die Polizistin Yelina Salas (Sofia Milos). Die beiden verbindet einerseits, dass Horatios Bruder mit Yelina verheiratet war, bevor er als krimineller Polizist enttarnt und getötet wurde. Außerdem entdecken die beiden, dass sie Gefühle füreinander entwickelt haben, worüber sie sich in dieser Szene lange unterhalten. Horatio öffnet sich Yelina und erzählt ihr, dass er aufgrund des Todes seines Bruders keinen Schritt auf sie zu machen kann und ihre Zuneigung füreinander nicht zu einer Beziehung führen kann, was beide ob der Ausweglosigkeit unglücklich zurücklässt. Die dargestellten Emotionen sind **Vertrauen** und **Traurigkeit**.

Sämtliche Ausschnitte waren nur wenige Sekunden kurz. Die Entscheidung dafür ist vor allem deshalb gefallen, weil so kein Befragungsteilnehmer durch das Darbieten eines Kontextes auf die gesuchte/n Emotion/en schließen konnte. Abgesehen davon sollte nur das Gesicht des Sprechers/der Sprecherin im Bild sein, was die Länge der Ausschnitte allein durch den ständigen Wechsel der Kameraperspektive in beiden Produktionen zeitlich

begrenzt. Außerdem erschweren sehr kurze Ausschnitte die Emotionserkennung für die Untersuchungsteilnehmer – was dazu führen sollte, dass bei der Auswertung stärker diskriminiert werden kann zwischen jenen Emotionen, die besonders gut und jenen, die besonders schwer zu erkennen sind.

3.4.2. Fragebogenerstellung

Der Fragebogen (siehe Anhang) war in verschiedene Abschnitte gegliedert. Vor dem Start der Befragung wurden die Teilnehmer instruiert – etwa, dass aus technischen Gründen eine Teilnahme lediglich mit dem „Mozilla Firefox“ als Webbrowser möglich war sowie die Bitte um vollständiges Ausfüllen und die Zusicherung der Anonymität.
Darauf folgte eine Stimmungserhebung zum sanften Einstieg. Es wurden jene acht Emotionen nach Plutchik (siehe 2.4.1.5.1) in einer unipolaren Ratingskala (von „Gar nicht“ bis „Sehr stark“) abgefragt, die später auch bei der Emotionserkennung per Audio- oder Videodatei zur Auswahl standen. So sollten die Befragungsteilnehmer sich bereits zu Beginn mit dieser speziellen Emotionsklassifizierung vertraut machen.

Auf die Stimmungserhebung folgte der erste Fragebogen zu den Themen Phantasieempathie und Empathie in realen Situationen, er umfasste insgesamt 21 Items. Diese entstammten der „E-Skala“, einem Fragebogen zur Erfassung von Empathie (siehe 3.3.2.1). Die Antwortmöglichkeiten waren ebenfalls in einer unipolaren Ratingskala (von „Trifft gar nicht zu“ bis „Trifft sehr zu“) angeordnet.

Der dritte Teil drehte sich um die Serienkenntnis der Teilnehmer, mit je zwei Fragen zu jeder TV-Serie. So wurde jeweils für „CSI: Miami“ und „Damages“ gefragt: „Schätzen Sie: Wie oft haben Sie diese Fernsehserie (auf Deutsch) INSGESAMT bereits gesehen?“ Die Antwortmöglichkeiten waren „Noch nie“, „1 Mal“, „2 bis 5 Mal“, „5 bis 10 Mal“ sowie „Öfter als 10 Mal“. Ebenfalls für beide Serien wurde gefragt: „Sehen Sie sich diese Serie REGELMÄSSIG (auf Deutsch) im Fernsehen an? Bzw. haben Sie die Serie REGELMÄSSIG (auf Deutsch) verfolgt, als sie im TV zu sehen war?“. Die Antworten reichten hier von „Verfolg(t)e ich nicht regelmäßig“ über „1 Mal pro Monat“ zu „2 bis 3 Mal pro Monat“ und „1 Mal pro Woche“ bis zu „Öfter als 1 Mal pro Woche“.

Der vierte Teil war der auditive, visuelle und audiovisuelle Fragebogenteil. Dabei standen am Anfang einige Instruktionen, die zum Beispiel darauf hinwiesen, dass die Audio-/Video-

Sequenzen sehr kurz seien und ein oder auch zwei Gefühl/e pro Sequenz angeklickt werden sollten. Jedes der sieben Beispiele war mit dem Hinweis darauf versehen, ob es sich um eine reine Audio-Datei ohne Bild, eine reine Video-Datei ohne Ton oder eine audiovisuelle Datei handelte. Zum einen, um Verwirrung zu vermeiden bei reinen Audio- und Video-Dateien und zum anderen, um Personen die trotz mehrmaligen Hinweises ihren Sound nicht aktiviert (z.B. die Boxen nicht angeschlossen) hatten, darauf aufmerksam zu machen. Unter jedem der sieben kurzen Ausschnitte stand die Frage „Was denken Sie: Welches Gefühl vermittelt der/die SprecherIn/SchauspielerIn?“. Zur Auswahl standen die acht Emotionen nach Plutchik, die bereits bei der Stimmungserhebung verwendet wurden. Es gab drei verschiedene Versionen des Fragebogens – nach dem Zufallsprinzip wurden die Befragungsteilnehmer einer von drei Gruppen zugeordnet, die jeweils eine Version jedes Ausschnitts sahen. Gruppe A sah das erste Beispiel als reine Video-Datei, Gruppe B dasselbe Beispiel als audiovisuelle Datei, Gruppe C erlebte lediglich die Audio-Datei (ohne Bild). Jede Gruppe erhielt reine Audio-, Video- sowie Audiovideo-Dateien. So war eine Randomisierung gegeben und eine unabhängige Stichprobe gewährleistet.

Im Anschluss waren in einem Persönlichkeitsfragebogen 56 Fragen zu beantworten. Diese Fragen entstammten dem Freiburger Persönlichkeitsinventar, kurz FPI-R (siehe 3.3.2.2). Zur Verwendung kamen die Items einzelner Skalen des FPI-R, und zwar diejenigen zur Bestimmung von Lebenszufriedenheit, sozialer Orientierung, Gehemmtheit, Extraversion und Emotionalität. Diese fünf Skalen beziehen sich auf die emotionale Welt der Befragten, auf ihren Umgang mit Gefühlen gegenüber sich selbst und anderen Personen und erschienen daher als geeignetes Mittel, die Auswirkung ebenjener Gefühlswelt auf das Erleben und Erkennen von Emotionen zu erforschen.
Anschließend wurden soziale Daten abgefragt: Geschlecht, Alter, höchster Schulabschluss, Berufstätigkeit, Familienstand sowie die Anzahl von Kindern und Geschwistern. So sollte geklärt werden, ob soziale Daten oder familiäre Umstände wie Kinder und Geschwister dazu beitragen, dass Menschen Emotionen besser erkennen können.
Ein Dankeschön für die Teilnahme und eine E-Mail-Adresse für Fragen und Rückmeldungen schlossen den Fragebogen ab.

3.4.3. Pretest des Fragebogens

Der Pretest des Fragebogens dauerte, vor allem aufgrund technischer Schwierigkeiten bei der Umsetzung, mehrere Wochen, dabei wurde er mehrmals von verschiedenen – insgesamt fünf

– Personen ausgefüllt. Dabei kam neben dem Freiburger Persönlichkeitsinventar auch der Trierer Persönlichkeitsfragebogen, TPF, zum Einsatz. Dabei wurden auch jene Szenen aus Fernsehserien ausgewählt, die sich für die Emotionserkennung als geeignet herausstellten.

3.4.4. Veränderungen am Fragebogen

Aufgrund des Pretests fiel die Entscheidung auf das Freiburger Persönlichkeitsinventar, da das dichotome Antwortformat und die Fragestellungen von den Pretest-Teilnehmern als klarer und verständlicher wahrgenommen wurden als beim TPF. Hier war es für die meisten Befragten schwierig, anhand der vorgegebenen Möglichkeiten „Immer“, „Oft“, „Manchmal“ und „Nie“ eindeutige Antworten auf Fragen wie z.B. „Ich bin … davon überzeugt, daß [sic] man mich sehr mögen kann.“ zu finden.
Nach weiteren Pretests mit der vollständigen Befragung inklusive FPI-R wurden geringfügige Änderungen vorgenommen, etwa das Einfügen von Hinweisen auf den Fortschritt beim Ausfüllen (bei jeder neuen Etappe wurde der Hinweis „Teil X von X“ – also z.B. „Teil 4 von 6“ angezeigt). Damit war gewährleistet, dass die Befragungsteilnehmer einen Fortschritt beobachten und auf ein Ziel – das Ende des Fragebogens – hinarbeiten konnten, was die Motivation steigern und kein Gefühl des Ausgeliefertseins aufkommen lassen sollte. Außerdem wurde der Hinweis, dass der Fragebogen lediglich mit dem „Mozilla Firefox“ zugänglich war, visuell noch stärker betont.

3.4.5. Untersuchungsdurchführung

Der Fragebogen wurde am 10. März 2009 auf meinem Unet-Account der Universität Wien (wo jedem Studenten Webspace zur Verfügung steht) online gestellt. Die letzten Fragebögen gelangten am 30. März 2009 ein, danach wurde der Fragebogen offline gestellt.

3.4.6. Fragebogenteilnehmer

Bei den Befragungsteilnehmern handelte sich um eine einfache Zufallsstichprobe, die allerdings mittels des Schneeballprinzips erreicht wurde. 104 gültige Fragebögen sind eingelangt – außerdem 74 nicht verwendbare. Diese hohe Zahl ist einerseits dadurch zu erklären, dass offenbar viele die Instruktionen am Beginn des Fragebogens nicht gelesen hatten und mit dem „Internet Explorer“ an der Umfrage teilnehmen wollten – wodurch sie zwar nicht weitergeleitet wurden, ihr Versuch aber dennoch vom System automatisch gespeichert wurde. Außerdem verweigerten etwa zehn Prozent der Teilnehmer mit nicht verwendbaren Fragebögen die Angabe von persönlichen Daten. Dazu kamen die im Internet

nicht erstaunlichen Abbrüche inmitten des Fragebogens – vermutlich aus Desinteresse oder Zeitmangel. Es ist allerdings nicht möglich nachzuvollziehen, wie viele Personen von den 74 nicht vollständig ausgefüllten Fragebögen die Untersuchung zu einem späteren Zeitpunkt noch einmal gestartet und vollständig zu Ende gebracht haben.

Der Fragebogen wurde per Email im Schneeballprinzip zunächst an Freunde verschickt (hauptsächliche Altersgruppe 20 bis 29 Jahre), außerdem an Freunde meiner Eltern (Altersgruppe 50 plus) – was die Aufteilung der Befragungsteilnehmer erklärt: Teilnehmer im Alter von 20 bis 29 Jahren sind mit 41,7 Prozent deutlich überrepräsentiert. Immerhin ist es aber gelungen, mit 32,1 Prozent der Teilnehmer viele Personen über 50 Jahren zur Teilnahme zu bewegen.

Altersgruppen

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	bis 19 Jahre	2	1,9	1,9	1,9
	20 bis 29 Jahre	43	41,3	41,7	43,7
	30 bis 39 Jahre	15	14,4	14,6	58,3
	40 bis 49 Jahre	10	9,6	9,7	68,0
	50 bis 59 Jahre	18	17,3	17,5	85,4
	60 bis 69 Jahre	14	13,5	13,6	99,0
	ab 70 Jahre	1	1,0	1,0	100,0
	Gesamt	103	99,0	100,0	
Fehlend	System	1	1,0		
Gesamt		104	100,0		

In Bezug auf den Beruf (Die Frage lautete: „Sind Sie berufstätig?“) schlug sich diese Überrepräsentation ebenfalls nieder – Studenten und solche, die auch arbeiten, machten gemeinsam 27,9 Prozent der Befragungsteilnehmer aus. Mit 51,9 Prozent stellten allerdings die Berufstätigen die mit Abstand größte Gruppe dar. Deutlich unterrepräsentiert waren arbeitslose und in Berufsausbildung befindliche Personen mit insgesamt nur 2 von 104 Befragten.

Beruf

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente

Gültig	arbeitslos/in Umbildung	1	1,0	1,0	1,0
	Hausfrau/Hausmann	6	5,8	5,8	6,7
	im Ruhestand	11	10,6	10,6	17,3
	in Berufsausbildung	1	1,0	1,0	18,3
	Ja	54	51,9	51,9	70,2
	Ja, und Student/in	13	12,5	12,5	82,7
	Schüler/in	2	1,9	1,9	84,6
	Student/in	16	15,4	15,4	100,0
	Gesamt	104	100,0	100,0	

Eine mögliche Verzerrung zeigt auch die Frage nach dem Schulabschluss, 46 Prozent hatten einen Universitäts- oder Fachhochschulabschluss, 37 Prozent Matura – insgesamt haben also 83 Prozent der Befragungsteilnehmer höhere Bildung genossen.

Schule

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Lehre, Fachschule	16	15,4	16,0	16,0
	Matura	37	35,6	37,0	53,0
	Pflichtschule	1	1,0	1,0	54,0
	Universität, Fachhochschule	46	44,2	46,0	100,0
	Gesamt	100	96,2	100,0	
Fehlend	.	4	3,8		
Gesamt		104	100,0		

Beim Familienstand sind die Gruppen derjenigen mit Freund/in oder Lebenspartner/in sowie der Verheirateten mit 35,6 bzw. 34,6 Prozent die mit Abstand größten. Auch diese Verteilung könnte eine Verzerrung darstellen – obwohl in vielen Befragungen und Bevölkerungsdaten Partnerschaften nicht abgefragt werden, sondern lediglich Ehen.

Familienstand

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Freund/in, Lebenspartner/in	37	35,6	35,6	35,6
	getrennt lebend/geschieden	5	4,8	4,8	40,4
	ledig	25	24,0	24,0	64,4

	verheiratet	36	34,6	34,6	99,0
	verwitwet	1	1,0	1,0	100,0
	Gesamt	104	100,0	100,0	

Immerhin 41,3 Prozent der Befragten hatten ein oder mehrere Kinder, allerdings erst ab 30 Jahren. Im Alter von 40 bis 59 Jahren hatten die meisten Befragten ein oder mehrere Kinder sowie sämtliche Untersuchungsteilnehmer ab 60 Jahren.

Kinder

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	ja	43	41,3	41,3	41,3
	nein	61	58,7	58,7	100,0
	Gesamt	104	100,0	100,0	

Altersgruppen * Kinder Kreuztabelle

			Kinder		Gesamt
			ja	nein	
Altersgruppen	bis 19 Jahre	Anzahl	0	2	2
		% innerhalb von Altersgruppen	,0%	100,0%	100,0%
	20 bis 29 Jahre	Anzahl	0	43	43
		% innerhalb von Altersgruppen	,0%	100,0%	100,0%
	30 bis 39 Jahre	Anzahl	3	12	15
		% innerhalb von Altersgruppen	20,0%	80,0%	100,0%
	40 bis 49 Jahre	Anzahl	8	2	10
		% innerhalb von Altersgruppen	80,0%	20,0%	100,0%
	50 bis 59 Jahre	Anzahl	16	2	18
		% innerhalb von Altersgruppen	88,9%	11,1%	100,0%
	60 bis 69 Jahre	Anzahl	14	0	14
		% innerhalb von Altersgruppen	100,0%	,0%	100,0%
	ab 70 Jahre	Anzahl	1	0	1
		% innerhalb von Altersgruppen	100,0%	,0%	100,0%
Gesamt		Anzahl	42	61	103
		% innerhalb von Altersgruppen	40,8%	59,2%	100,0%

3.5. Ergebnisse

Forschungsfrage 1: Können Zuhörer anhand eines Serien-Ausschnitts, der aus einer stimmlichen Äußerung von nur wenigen Sekunden besteht, erkennen, welche Emotion/en vom Synchronsprecher vermittelt wird/werden?

Hypothese 1.0: Zuhörer können anhand eines kurzen Ausschnitts nicht erkennen, welche Emotion/en vermittelt wird/werden.

Hypothese 1.1: Zuhörer können anhand eines kurzen Ausschnitts erkennen, welche Emotion/en vermittelt wird/werden.

Um diese Frage zu beantworten, wurde für jede Gruppe der Erwartungswert mittels Wahrscheinlichkeitsfunktion der hypergeometrischen Verteilung berechnet. Dieser Wert sagt aus, wie viel Punkte (in Prozent) eine Person erreichen würde, die die Emotionen nicht aktiv erkennt sondern lediglich (sozusagen blindlings) rät. Für die Gruppen A und C ergab sich ein Erwartungswert von E = 25 %, für die Gruppe B von E = 21 %. Im Vergleich mit den beobachteten Prozentwerten mit einem Mittelwert von 39,50 % wird anhand des Wilcoxon-Tests deutlich, dass ein signifikanter Unterschied (p = 0,000) besteht. Die Fragebogenteilnehmer haben die Emotionen im Durchschnitt also signifikant besser erkannt als Personen, die per Zufall Antworten auswählen würden – diese also nicht unterscheiden und erkennen könnten.

Deskriptive Statistiken

	N	Mittelwert	Standardabweichung	Minimum	Maximum
Alle Videos - Auditiv richtig erkannt in Prozent	104	39,5032	25,01335	,00	100,00
Auditiv richtig erkannt - Erwartungswert	104	23,7308	1,87073	21,00	25,00

Ränge

		N	Mittlerer Rang	Rangsumme
Auditiv richtig erkannt - Erwartungswert - Alle Videos - Auditiv richtig erkannt in Prozent	Negative Ränge	64[a]	50,13	3208,00
	Positive Ränge	24[b]	29,50	708,00
	Bindungen	16[c]		
	Gesamt	104		

a. Auditiv richtig erkannt - Erwartungswert < Alle Videos - Auditiv richtig erkannt in Prozent

b. Auditiv richtig erkannt - Erwartungswert > Alle Videos - Auditiv richtig erkannt in Prozent

c. Auditiv richtig erkannt - Erwartungswert = Alle Videos - Auditiv richtig erkannt in Prozent

Statistik für Test[b]

	Auditiv richtig erkannt - Erwartungswert - Alle Videos - Auditiv richtig erkannt in Prozent
Z	-5,252[a]
Asymptotische Signifikanz (2-seitig)	,000

a. Basiert auf positiven Rängen.

b. Wilcoxon-Test

Hypothese 1.0 ist demnach falsifiziert, Hypothese 1.1 darf hingegen als bestätigt angenommen werden.

Die Punkte, die den Mittelwert von 39,50 ergaben, wurden folgendermaßen berechnet: Bei sechs von sieben Sequenzen waren zwei Emotionen (statt nur einer) gesucht, was den Befragten allerdings nicht konkret angegeben wurde – sie wurden lediglich am Beginn dieses Fragebogen-Teils darauf hingewiesen, dass sie sowohl eine als auch zwei Emotionen pro Sequenz angeben könnten. Erkannte eine Person eine von zwei Emotionen, erhielt sie einen Punkt, zwei von zwei Emotionen ergaben folglich zwei Punkte. Bei der Sequenz mit nur einer gesuchten Emotion (Angst) gab es nur einen Punkt für eine richtige Antwort. Falsche Antworten wurden in keinem Fall mit Punkteabzug geahndet. Die Summe der erreichten Punkte wurde durch die maximal zu erreichende Punkteanzahl dividiert und mit 100 multipliziert, was den individuellen Prozentwert an richtig erkannten Emotionen ergab.

Alle Videos - Auditiv richtig erkannt in Prozent

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	,00	12	11,5	11,5	11,5
	16,67	12	11,5	11,5	23,1
	25,00	16	15,4	15,4	38,5
	33,33	19	18,3	18,3	56,7
	50,00	16	15,4	15,4	72,1
	66,67	24	23,1	23,1	95,2
	75,00	1	1,0	1,0	96,2

	100,00	4	3,8	3,8	100,0
	Gesamt	104	100,0	100,0	

Wird außer Acht gelassen, dass bei sechs von sieben Sequenzen zwei Emotionen gesucht waren, rechnet man also bereits eine richtig erkannte Emotion pro Ausschnitt als 100 statt 50 Prozent der möglichen Punkte dieser Sequenz, verändert sich die Statistik folgendermaßen:

Statistiken

1 Emotion pro Sequenz auditiv richtig erkannt in Prozent

N	Gültig	104
	Fehlend	0
Mittelwert		68,5897

1 Emotion pro Sequenz auditiv richtig erkannt in Prozent

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	,00	12	11,5	11,5	11,5
	33,33	12	11,5	11,5	23,1
	50,00	21	20,2	20,2	43,3
	66,67	14	13,5	13,5	56,7
	100,00	40	38,5	38,5	95,2
	150,00	5	4,8	4,8	100,0
	Gesamt	104	100,0	100,0	

Forschungsfrage 2: Welche Emotionen werden von Zuhörern besonders klar erkannt und welche sind schwer zu identifizieren?

Hypothese 2.0: Emotionen werden unabhängig von ihrer Valenz in ähnlichem Ausmaß erkannt.

Hypothese 2.1: Stark negative Emotionen werden im Allgemeinen besser erkannt als positive und neutrale Emotionen.

Hypothese 2.2: Positive Emotionen werden im Allgemeinen besser erkannt als neutrale Emotionen.

Aufgrund des komplexen Sampledesigns war es mit SPSS nicht möglich, eine Signifikanz zu errechnen. Die errechneten Prozentzahlen „[Emotion] auditiv richtig erkannt in Prozent“

ergeben sich für jeden Befragungsteilnehmer aus den Emotionen, die er anhand einer reinen Audio-Datei gehört und wie viele Punkte er dabei von den möglichen Punkten erreicht hatte.

Statistiken

		Traurigkeit auditiv richtig erkannt in Prozent	Ekel auditiv richtig erkannt in Prozent	Überraschung auditiv richtig erkannt in Prozent	Vertrauen auditiv richtig erkannt in Prozent
N	Gültig	71	36	104	71
	Fehlend	33	68	0	33
Mittelwert		26,7606	37,5000	25,9615	28,1690
Modus		,00	50,00	,00	,00
Standardabweichung		44,58618	27,71024	44,05467	45,30247

Statistiken

		Angst auditiv richtig erkannt in Prozent	Wut auditiv richtig erkannt in Prozent	Neugierde auditiv richtig erkannt in Prozent	Freude auditiv richtig erkannt in Prozent
N	Gültig	33	36	33	35
	Fehlend	71	68	71	69
Mittelwert		93,9394	52,7778	39,3939	,0000
Modus		100,00	100,00	,00	,00
Standardabweichung		24,23058	50,63094	49,61977	,00000

Allerdings sprechen die Ergebnisse auch ohne Signifikanztest eine deutliche Sprache. So wird aus den Häufigkeitstabellen klar, dass negative Emotionen deutlich besser erkannt wurden als positive oder neutrale Emotionen.

Angst wurde von 93,9 Prozent der Befragten anhand eines rein auditiven Ausschnitts richtig erkannt (Angst war nur einmal zu hören, daher 0 oder 100 Prozent). Wut erkannten 52,8 Prozent der Befragten (anhand eines Ausschnitts) richtig. Ekel wurde (bei zwei Chancen, da zwei Ausschnitte Ekel beinhalten) von immerhin 63,9 Prozent der Befragten bei einer Gelegenheit richtig erkannt – der Mittelwert liegt hier dennoch bei lediglich 37,5 Prozent. Deutlich schlechter fiel die Erkennungsquote bei Traurigkeit aus – 26,8 Prozent der Befragten erkannten diese Emotion.

Bei den neutralen Emotionen zeigte sich ein geteiltes Bild: Bei Überraschung lagen nur 26 Prozent richtig, Neugierde wurde mit 39,4 Prozent richtigen Antworten deutlich besser erkannt.

Etwas besser als Traurigkeit und Überraschung wurde die positive Emotion Vertrauen erkannt – mit 28,2 Prozent richtigen Antworten bei einem Ausschnitt. Mit Abstand am schlechtesten wurde anhand eines rein auditiven Ausschnitts mit Freude die positivste Emotion erkannt – kein einziger der Befragten konnte sie identifizieren.

Anhand dieser Ergebnisse darf die Nullhypothese als nicht korrekt eingestuft werden. Hypothese 2.1 hingegen darf als verifiziert gelten, da die Gefühle mit der höchsten negativen Intensität und Valenz – Angst und Wut – mit Abstand besser erkannt wurden als neutrale und positive Emotionen. Hypothese 2.2 hingegen hat sich nicht bestätigt – besonders die positivste Emotion, die Freude, wurde anhand reiner Audio-Dateien gar nicht erkannt und damit deutlich schlechter als Emotionen mit geringerer Valenz und Intensität.

Forschungsfrage 3: Wodurch unterscheiden sich die Ergebnisse zwischen Zuhörern und Zusehern, denen nur das zur Stimme gehörige Bild ohne Ton vorgelegt wird?
Hypothese 3.0: Emotionen werden im Allgemeinen gleich gut erkannt**.**
Hypothese 3.1: Zuhörer erkennen Emotionen besser als die Zuseher.

Da die Daten nicht normalverteilt sind (siehe Signifikanz von p=0,007 bzw. p=0,000 des Kolmogorov-Smirnov-Tests), wurden die Ergebnisse der Emotionserkennung bei reinen Audio-Dateien sowie reinen Video-Dateien dem Friedman-Test unterzogen. Die Prozentangaben wurden aus den je nach Datei-Art richtig erkannten Emotionen sowie den maximal zu erreichenden Punkten berechnet.

Kolmogorov-Smirnov-Anpassungstest

		Alle Videos - Auditiv richtig erkannt in Prozent	Alle Videos - Visuell richtig erkannt in Prozent
N		104	104
Parameter der Normalverteilung[a.,b]	Mittelwert	39,5032	48,5096
	Standardabweichung	25,01335	16,55939
Extremste Differenzen	Absolut	,165	,224
	Positiv	,165	,224
	Negativ	-,140	-,199
Kolmogorov-Smirnov-Z		1,680	2,282
Asymptotische Signifikanz (2-seitig)		,007	,000

a. Die zu testende Verteilung ist eine Normalverteilung.

b. Aus den Daten berechnet.

Ränge

		N	Mittlerer Rang	Rangsumme
Alle Videos - Visuell richtig erkannt in Prozent - Alle Videos - Auditiv richtig erkannt in Prozent	Negative Ränge	35[a]	47,49	1662,00
	Positive Ränge	64[b]	51,38	3288,00
	Bindungen	5[c]		
	Gesamt	104		

a. Alle Videos - Visuell richtig erkannt in Prozent < Alle Videos - Auditiv richtig erkannt in Prozent

b. Alle Videos - Visuell richtig erkannt in Prozent > Alle Videos - Auditiv richtig erkannt in Prozent

c. Alle Videos - Visuell richtig erkannt in Prozent = Alle Videos - Auditiv richtig erkannt in Prozent

Statistik für Test[b]

	Alle Videos - Visuell richtig erkannt in Prozent - Alle Videos - Auditiv richtig erkannt in Prozent
Z	-2,846[a]
Asymptotische Signifikanz (2-seitig)	,004

a. Basiert auf negativen Rängen.

b. Wilcoxon-Test

Der Wilcoxon-Test zeigte mit einer Signifikanz von p = 0,004, dass die Nullhypothese auszuschließen ist. Genaueren Aufschluss gaben die gruppierten Ergebnisse der Emotionserkennung bei reinen Audio- und Video-Dateien:

Statistiken

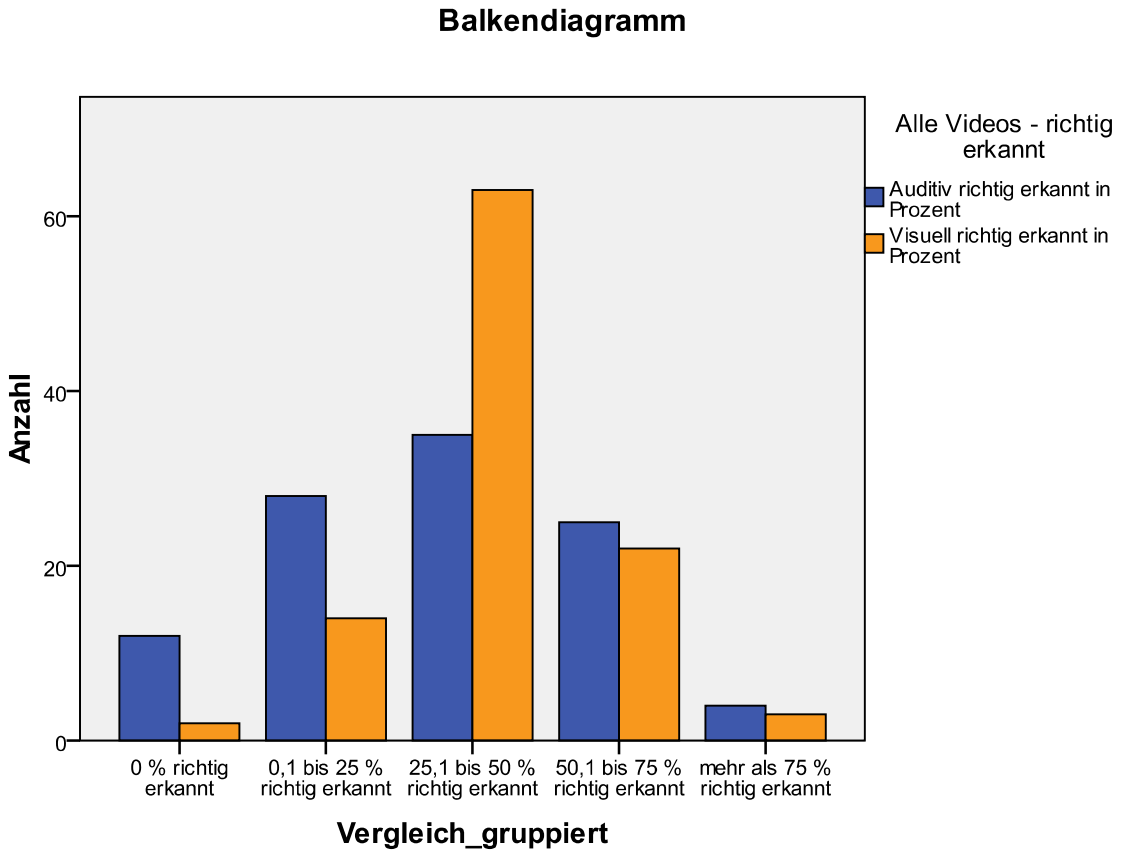
		Alle Videos - Auditiv richtig erkannt - Gruppen Prozent	Alle Videos - Visuell richtig erkannt - Gruppen Prozent
N	Gültig	104	104
	Fehlend	0	0

Alle Videos - Auditiv richtig erkannt - Gruppen Prozent

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	0 % richtig erkannt	12	11,5	11,5	11,5
	0,1 bis 25 % richtig erkannt	28	26,9	26,9	38,5
	25,1 bis 50 % richtig erkannt	35	33,7	33,7	72,1
	50,1 bis 75 % richtig erkannt	25	24,0	24,0	96,2
	mehr als 75 % richtig erkannt	4	3,8	3,8	100,0
	Gesamt	104	100,0	100,0	

Alle Videos - Visuell richtig erkannt - Gruppen Prozent

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	0 % richtig erkannt	2	1,9	1,9	1,9
	0,1 bis 25 % richtig erkannt	14	13,5	13,5	15,4
	25,1 bis 50 % richtig erkannt	63	60,6	60,6	76,0
	50,1 bis 75 % richtig erkannt	22	21,2	21,2	97,1
	mehr als 75 % richtig erkannt	3	2,9	2,9	100,0
	Gesamt	104	100,0	100,0	



Bereits der Vergleich der Mittelwerte – Audio: 39,5 Prozent, Video 48,5 Prozent – zeigt, dass die Befragungsteilnehmer größere Schwierigkeiten mit der Emotionserkennung anhand von Audio-Dateien als Video-Dateien hatten. Die gruppierten Ergebnisse geben Aufschluss über die Verteilung innerhalb der Gruppen: 11,5 Prozent der Befragten konnten anhand der Audio-Ausschnitte keine Emotion richtig zuordnen, bei den reinen Video-Dateien lag dieser Wert bei lediglich 1,9 Prozent. Bis zu 25 Prozent der zu erkennenden Emotionen wurden von 26,9 Prozent der Zuhörer identifiziert, was den kumulierten Prozentwert auf 38,5 Prozent steigert. Dieser Wert liegt bei der Gruppe der Zuseher ohne Ton bei lediglich 15,4 Prozent. Auch die Gruppe derjenigen, die 25,1 bis 50 Prozent der Emotionen richtig erkannt haben, lag bei den Audio-Dateien mit 33,7 Prozent weit hinter den Video-Ausschnitten mit 60,6 Prozent. Die Gruppe derjenigen, die die Hälfte bis zu sämtlichen dargestellten Emotionen richtig erfassen, war hingegen mit insgesamt 27,8 Prozent bei den reinen Audio-Dateien geringfügig größer als bei den Videos mit insgesamt 24,1 Prozent.

Hypothese 3.1 darf aufgrund dieser Ergebnisse als widerlegt angesehen werden, im Gegenteil werden Emotionen von reinen Zusehern besser erkannt als von Zuhörern.

Hypothese 3.2: Zuhörer können stark negative Emotionen besser erkennen als Zuseher.
Hypothese 3.3: Zuhörer können stark positive Emotionen besser erkennen als Zuseher.

Um die Hypothesen 3.2 und 3.3 zu überprüfen, konnte wiederum aufgrund der Datenstruktur nur bei den Emotionen Traurigkeit sowie Vertrauen ein Signifikanztest durchgeführt werden. Bei beiden Emotionen zeigte sich beim Kolmogorov-Smirnov-Test eine Signifikanz (p=0,000), weshalb in weiterer Folge der Wilcoxon-Test zur Anwendung kam.

Kolmogorov-Smirnov-Anpassungstest

		Traurigkeit auditiv richtig erkannt in Prozent	Traurigkeit visuell richtig erkannt in Prozent
N		71	69
Parameter der Normalverteilung[a,,b]	Mittelwert	26,7606	60,8696
	Standardabweichung	44,58618	49,16177
Extremste Differenzen	Absolut	,458	,396
	Positiv	,458	,283
	Negativ	-,274	-,396
Kolmogorov-Smirnov-Z		3,861	3,287
Asymptotische Signifikanz (2-seitig)		,000	,000

a. Die zu testende Verteilung ist eine Normalverteilung.

b. Aus den Daten berechnet.

Kolmogorov-Smirnov-Anpassungstest

		Vertrauen auditiv richtig erkannt in Prozent	Vertrauen visuell richtig erkannt in Prozent
N		71	69
Parameter der Normalverteilung[a,,b]	Mittelwert	28,1690	44,9275
	Standardabweichung	45,30247	50,10645
Extremste Differenzen	Absolut	,451	,366
	Positiv	,451	,366
	Negativ	-,267	-,313
Kolmogorov-Smirnov-Z		3,802	3,038
Asymptotische Signifikanz (2-seitig)		,000	,000

a. Die zu testende Verteilung ist eine Normalverteilung.

b. Aus den Daten berechnet.

- **Traurigkeit:**

Ränge

		N	Mittlerer Rang	Rangsumme
Traurigkeit visuell richtig erkannt in Prozent - Traurigkeit auditiv richtig erkannt in Prozent	Negative Ränge	1[a]	13,00	13,00
	Positive Ränge	24[b]	13,00	312,00
	Bindungen	11[c]		
	Gesamt	36		

a. Traurigkeit visuell richtig erkannt in Prozent < Traurigkeit auditiv richtig erkannt in Prozent

b. Traurigkeit visuell richtig erkannt in Prozent > Traurigkeit auditiv richtig erkannt in Prozent

c. Traurigkeit visuell richtig erkannt in Prozent = Traurigkeit auditiv richtig erkannt in Prozent

Statistik für Test[b]

	Traurigkeit visuell richtig erkannt in Prozent - Traurigkeit auditiv richtig erkannt in Prozent
Z	-4,600[a]
Asymptotische Signifikanz (2-seitig)	,000

a. Basiert auf negativen Rängen.

b. Wilcoxon-Test

Die Bedeutung der Signifikanz des Wilcoxon-Tests (p=0,000) wird anhand der Häufigkeitstabelle deutlich: Während Traurigkeit lediglich von 26,8 Prozent der reinen Zuhörer erkannt wurde, konnten 60,9 Prozent die Emotion anhand der Video-Datei richtig angeben.

Traurigkeit ist folglich in einem Video ohne Ton wesentlich deutlicher zu erkennen als anhand einer stimmlichen Äußerung.

Statistiken

		Traurigkeit auditiv richtig erkannt in Prozent	Traurigkeit visuell richtig erkannt in Prozent
N	Gültig	71	69
	Fehlend	33	35

Mittelwert	26,7606	60,8696
Modus	,00	100,00

- **Vertrauen:**

Ränge

		N	Mittlerer Rang	Rangsumme
Vertrauen visuell richtig erkannt in Prozent - Vertrauen auditiv richtig erkannt in Prozent	Negative Ränge	3[a]	4,50	13,50
	Positive Ränge	5[b]	4,50	22,50
	Bindungen	28[c]		
	Gesamt	36		

a. Vertrauen visuell richtig erkannt in Prozent < Vertrauen auditiv richtig erkannt in Prozent

b. Vertrauen visuell richtig erkannt in Prozent > Vertrauen auditiv richtig erkannt in Prozent

c. Vertrauen visuell richtig erkannt in Prozent = Vertrauen auditiv richtig erkannt in Prozent

Statistik für Test[b]

	Vertrauen visuell richtig erkannt in Prozent - Vertrauen auditiv richtig erkannt in Prozent
Z	-,707[a]
Asymptotische Signifikanz (2-seitig)	,480

a. Basiert auf negativen Rängen.

b. Wilcoxon-Test

Der Wilcoxon-Test zeigt bei der Emotion Vertrauen keine Signifikanz (p=0,480). Anhand der Audio-Datei wurde Vertrauen von 28,2 Prozent der Befragungsteilnehmer korrekt identifiziert, bei der reinen Video-Datei lag dieser Anteil bei 44,9 Prozent.
Vertrauen ist folglich anhand einer Stimme nicht besser zu erkennen als anhand eines Videoausschnitts ohne Ton – der Prozentanteil des richtigen Erkennens liegt bei der visuellen Datei sogar (nicht signifikant) über dem der Audiodatei.

Statistiken

	Vertrauen auditiv richtig erkannt in Prozent	Vertrauen visuell richtig erkannt in Prozent

N	Gültig	71	69
	Fehlend	33	35
Mittelwert		28,1690	44,9275
Modus		,00	,00

- **Ekel:**

Statistiken

		Ekel auditiv richtig erkannt in Prozent	Ekel visuell richtig erkannt in Prozent
N	Gültig	36	68
	Fehlend	68	36
Mittelwert		37,5000	5,8824
Modus		50,00	,00
Standardabweichung		27,71024	23,70435

Ekel auditiv richtig erkannt in Prozent

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	,00	11	10,6	30,6	30,6
	50,00	23	22,1	63,9	94,4
	100,00	2	1,9	5,6	100,0
	Gesamt	36	34,6	100,0	
Fehlend	System	68	65,4		
Gesamt		104	100,0		

Ekel visuell richtig erkannt in Prozent

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	,00	64	61,5	94,1	94,1
	100,00	4	3,8	5,9	100,0
	Gesamt	68	65,4	100,0	
Fehlend	System	36	34,6		
Gesamt		104	100,0		

Bei der Emotion Ekel war aufgrund des Sampledesigns kein Signifikanztest möglich. Außerdem erschwert die unterschiedliche Chancenaufteilung (Ekel war zwei Mal als Audio-Datei, jedoch nur ein Mal als reine Video-Datei im Fragebogen) die Auswertung. Dennoch

zeigen die Häufigkeitstabellen eindeutig, dass Ekel anhand der Stimme wesentlich deutlicher erkannt wurde als mit dem Sehsinn. 30, 6 Prozent der Audio-Hörer erkannten keines von beiden Malen die Emotion Ekel, wohingegen 94,1 Prozent der visuellen Beobachter Ekel nicht identifizierten. Auch Mittel- und Modalwert unterstützen diese Beobachtung: Ersterer liegt bei der Audio-Erkennung bei 37,5 Prozent, bei der Video-Datei jedoch nur bei 5,88 Prozent.
Ekel ist demnach anhand der Stimme wesentlich deutlicher zu erkennen als anhand einer reinen Audio-Datei.

- **Angst:**

Statistiken

		Angst auditiv richtig erkannt in Prozent	Angst visuell richtig erkannt in Prozent
N	Gültig	33	35
	Fehlend	71	69
Mittelwert		93,9394	94,2857
Modus		100,00	100,00
Standardabweichung		24,23058	23,55041

Trotz fehlenden Signifikanztests ist aus der Statistik zweifelsfrei zu entnehmen, dass Angst anhand einer Audio-Datei mit nahezu gleich hoher Wahrscheinlichkeit richtig erkannt wurde wie anhand einer reinen Video-Datei.
Angst wird folglich sowohl durch eine Stimme wie auch durch einen Video-Ausschnitt gleich deutlich wahrgenommen und identifiziert.

- **Wut:**

Statistiken

		Wut auditiv richtig erkannt in Prozent	Wut visuell richtig erkannt in Prozent
N	Gültig	36	33
	Fehlend	68	71
Mittelwert		52,7778	78,7879
Modus		100,00	100,00
Standardabweichung		50,63094	41,51488

Aufgrund des nicht durchführbaren Signifikanztests ist eine gesicherte Aussage darüber, ob der Unterschied in der Erkennung der Emotion Wut zwischen der auditiven und der visuellen Teilnehmergruppe signifikant ist, nicht zu treffen. Aus den Häufigkeitstabellen lässt sich jedoch ablesen, dass Wut anhand einer Video-Datei von 78,8 Prozent der Befragten korrekt erkannt wurde – der Wert liegt genau 26 Prozentpunkte über dem der reinen Audio-Hörer. Diese erkannten Wut mit einer Genauigkeit von 52,8 Prozent.
Es ist folglich davon auszugehen, dass Wut anhand Video-Datei deutlich besser identifiziert wird als anhand einer stimmlichen Äußerung.

- **Neugierde:**

Statistiken

		Neugierde auditiv richtig erkannt in Prozent	Neugierde visuell richtig erkannt in Prozent
N	Gültig	33	35
	Fehlend	71	69
Mittelwert		39,3939	40,0000
Modus		,00	,00
Standardabweichung		49,61977	49,70501

Trotz fehlender Signifikanzprüfung lässt sich anhand der Häufigkeitstabellen eindeutig feststellen, dass Neugierde von Zuhörern wie von Zusehern beinahe exakt gleich gut bzw. schlecht identifiziert wird. 39,4 Prozent der Befragten konnten diese Emotion anhand einer Audio-Datei zuordnen, 40 Prozent anhand einer reinen Video-Datei.
Neugierde wird demnach gleich gut anhand einer stimmlichen Äußerung wie eines Videos ohne Ton erkannt.

- **Überraschung:**

Statistiken

		Überraschung auditiv richtig erkannt in Prozent	Überraschung visuell richtig erkannt in Prozent
N	Gültig	104	71
	Fehlend	0	33
Mittelwert		25,9615	25,3521
Modus		,00	,00

Statistiken

		Überraschung auditiv richtig erkannt in Prozent	Überraschung visuell richtig erkannt in Prozent
N	Gültig	104	71
	Fehlend	0	33
Mittelwert		25,9615	25,3521
Modus		,00	,00
Standardabweichung		44,05467	33,67199

Die Emotion Überraschung ist aufgrund der Gruppeneinteilung im Fragebogen schwer zu vergleichen, da sie zwei Mal als Video-Datei, jedoch nur einmal als Audio-Datei eingebunden war. Aus dem nahezu identischen Mittelwert von 25,96 bzw. 25,35 Prozent lässt sich jedoch ablesen, dass die Emotion sowohl auditiv als auch visuell gleich gut bzw. schlecht erkannt wurde.
Überraschung ist demnach weder allein anhand einer Stimme noch eines Videoausschnitts ohne Ton besser zu erkennen.

- **Freude:**

Statistiken

		Freude auditiv richtig erkannt in Prozent	Freude visuell richtig erkannt in Prozent
N	Gültig	35	36
	Fehlend	69	68
Mittelwert		,0000	80,5556
Modus		,00	100,00
Standardabweichung		,00000	40,13865

Die positivste Emotion, Freude, wurde anhand der Audio-Datei von keinem Fragebogenteilnehmer erkannt. Beinahe das Gegenteil ist bei der reinen Video-Datei der Fall, hier konnten 80,6 Prozent der Befragten die Emotion korrekt zuordnen.
Freude ist folglich anhand einer Video-Datei ohne Ton mit großem Abstand wesentlich besser zu identifizieren als durch eine stimmliche Äußerung.

Sowohl Hypothese 3.2 als auch 3.3 sind demnach widerlegt. Ekel ist die einzige negative Emotion, die anhand der Stimme wesentlich besser erkannt wird als anhand eines Videos ohne Ton. Angst wird bei stimmlichen Äußerungen sowie Videos ohne Ton gleichermaßen richtig zugeordnet. Traurigkeit und Wut werden von Zusehern wesentlich besser identifiziert als von Zuhörern. Die positiven Emotionen Vertrauen und Freude sind durch Videos ohne Ton deutlicher zu erkennen als durch stimmliche Äußerungen – vor allem bei Freude zeigen sich eklatante Unterschiede. Emotionen mit eher neutraler Valenz und niedriger Intensität wie Überraschung und Neugierde werden im Allgemeinen gleich gut erkannt.

Aufgrund dieser Ergebnisse lässt sich folgende neue Hypothese generieren:
Hypothese 3.4: Zuseher erkennen Emotionen – unabhängig von ihrer Valenz – im Allgemeinen besser als Zuhörer.

Forschungsfrage 4: Welche Unterschiede zeigen sich in den Ergebnissen zwischen reinen Zuhörern, reinen Zusehern sowie jenen Personen, die die betreffenden Ausschnitte mit Bild und Ton sehen?
Hypothese 4.0: Emotionen werden im Allgemeinen gleich gut erkannt.
Hypothese 4.1: Wer den Ausschnitt audiovisuell erlebt, kann Emotion/en besser zuordnen als reine Zuhörer und Zuseher.

Da die Daten keine Normalverteilung aufwiesen (siehe Kolmogorov-Smirnov-Test Forschungsfrage 3), wurde wiederum der Friedman-Test angewendet, der eine Signifikanz von p=0,005 aufwies.

Ränge

	Mittlerer Rang
Alle Videos - Auditiv richtig erkannt in Prozent	1,75
Alle Videos - Visuell richtig erkannt in Prozent	2,11
Alle Videos - Audiovisuell richtig erkannt in Prozent	2,14

Statistik für Test[a]

N	104
Chi-Quadrat	10,784
df	2

Asymptotische Signifikanz	,005

a. Friedman-Test

Diese Signifikanz wird anhand der Statistiken sichtbar. Der Mittelwert bei der richtigen Erkennung von Audio-Dateien lag bei 39,50 Prozent, bei visuellen Dateien bei 48,50 Prozent und bei audiovisuellen Dateien bei 49,92 Prozent. Der kumulierte Prozentwert jener Befragungsteilnehmer, die bis zu 50 Prozent der Emotionen richtig zuordneten, lag bei den audiovisuellen Dateien bei 56,7 Prozent – folglich haben 43,3 Prozent der Befragten anhand von Videos mit Ton mehr als die Hälfte der Emotionen korrekt identifiziert. Bei den Audio-Dateien lag dieser Wert bei lediglich 27,8 Prozent, bei den Video-Dateien bei 24,1 Prozent.

Statistiken

		Alle Videos - Auditiv richtig erkannt in Prozent	Alle Videos - Visuell richtig erkannt in Prozent	Alle Videos - Audiovisuell richtig erkannt in Prozent
N	Gültig	104	104	104
	Fehlend	0	0	0
Mittelwert		39,5032	48,5096	49,9199
Modus		66,67	50,00	66,67
Standardabweichung		25,01335	16,55939	26,28815

Alle Videos - Auditiv richtig erkannt in Prozent

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	0 % richtig erkannt	12	11,5	11,5	11,5
	0,1 bis 25 % richtig erkannt	28	26,9	26,9	38,5
	25,1 bis 50 % richtig erkannt	35	33,7	33,7	72,1
	50,1 bis 75 % richtig erkannt	25	24,0	24,0	96,2
	mehr als 75 % richtig erkannt	4	3,8	3,8	100,0
	Gesamt	104	100,0	100,0	

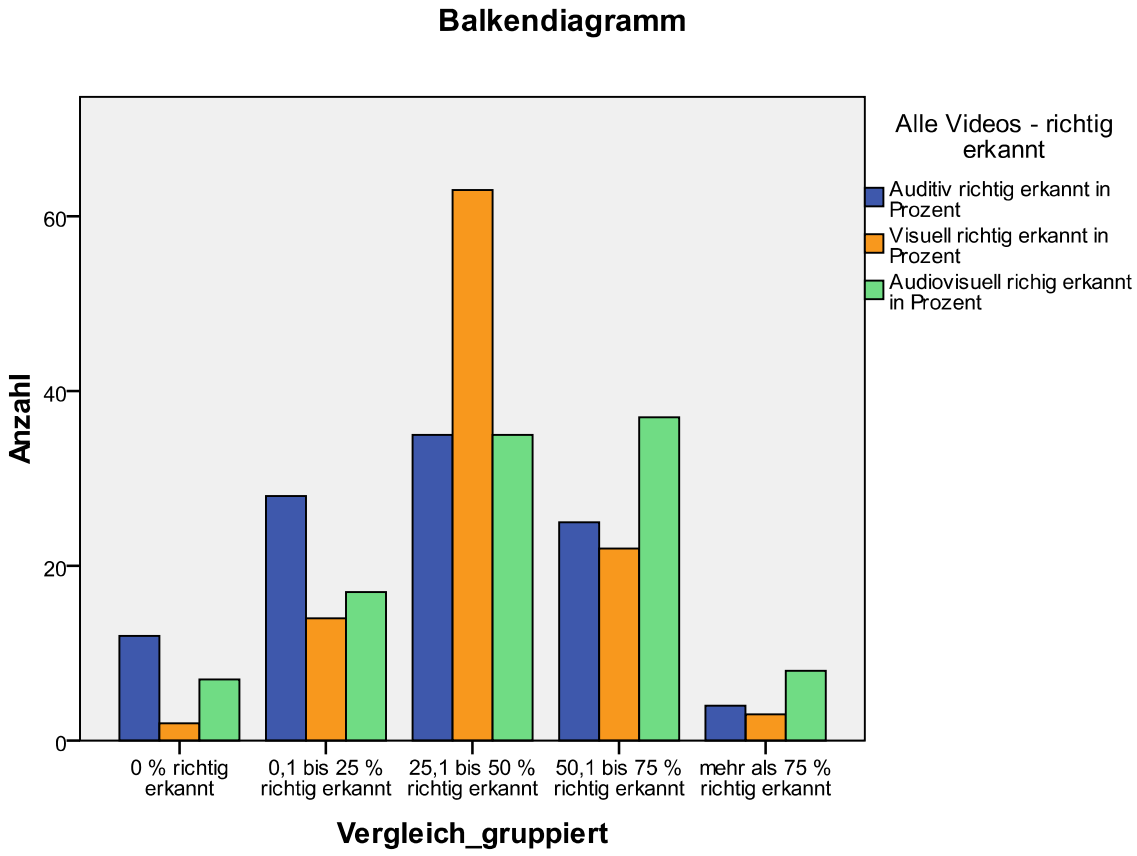
Alle Videos - Visuell richtig erkannt in Prozent

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	0 % richtig erkannt	2	1,9	1,9	1,9
	0,1 bis 25 % richtig erkannt	14	13,5	13,5	15,4

	25,1 bis 50 % richtig erkannt	63	60,6	60,6	76,0
	50,1 bis 75 % richtig erkannt	22	21,2	21,2	97,1
	mehr als 75 % richtig erkannt	3	2,9	2,9	100,0
	Gesamt	104	100,0	100,0	

Alle Videos - Audiovisuell richtig erkannt in Prozent

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	0 % richtig erkannt	7	6,7	6,7	6,7
	0,1 bis 25 % richtig erkannt	17	16,3	16,3	23,1
	25,1 bis 50 % richtig erkannt	35	33,7	33,7	56,7
	50,1 bis 75 % richtig erkannt	37	35,6	35,6	92,3
	mehr als 75 % richtig erkannt	8	7,7	7,7	100,0
	Gesamt	104	100,0	100,0	



Hypothese 4.0 ist demnach auszuschließen.

Um Hypothese 4.1 zu überprüfen, kamen zwei gesonderte Wilcoxon-Tests zum Einsatz. Hoch signifikant (p=0,017) zeigte sich dabei der Unterschied zwischen den reinen Audio-Dateien und den audiovisuellen Dateien. Wie aus obigen Tabellen, etwa dem Mittelwert von 39,50 Prozent (Audio) bzw. 49,92 Prozent (audiovisuell), zu entnehmen, wurden Emotionen anhand audiovisueller Dateien wesentlich besser erkannt als lediglich anhand der Stimmaufnahmen.

Ränge

		N	Mittlerer Rang	Rangsumme
Alle Videos - Audiovisuell richtig erkannt in Prozent - Alle Videos - Auditiv richtig erkannt in Prozent	Negative Ränge	34[a]	44,93	1527,50
	Positive Ränge	58[b]	47,42	2750,50
	Bindungen	12[c]		
	Gesamt	104		

a. Alle Videos - Audiovisuell richtig erkannt in Prozent < Alle Videos - Auditiv richtig erkannt in Prozent

b. Alle Videos - Audiovisuell richtig erkannt in Prozent > Alle Videos - Auditiv richtig erkannt in Prozent

c. Alle Videos - Audiovisuell richtig erkannt in Prozent = Alle Videos - Auditiv richtig erkannt in Prozent

Statistik für Test[b]

	Alle Videos - Audiovisuell richtig erkannt in Prozent - Alle Videos - Auditiv richtig erkannt in Prozent
Z	-2,391[a]
Asymptotische Signifikanz (2-seitig)	,017

a. Basiert auf negativen Rängen.

b. Wilcoxon-Test

Der Wilcoxon-Test zum Vergleich der Ergebnisse der reinen Video- sowie Audiovideo-Ausschnitte zeigte hingegen, dass die Unterschiede zwischen diesen nicht signifikant sind (p=0,544). Hypothese 4.1 ist demnach ebenfalls nicht korrekt. Wer einen Ausschnitt audiovisuell erlebt, kann Emotionen mit ähnlicher Genauigkeit erkennen wie jemand, der denselben Ausschnitt ohne Ton zu sehen bekommt. Beide Rezeptionsformen lassen Emotionen jedoch signifikant (p=0,005) besser erkennen als der stimmliche Ausdruck allein.

Ränge

		N	Mittlerer Rang	Rangsumme
Alle Videos - Audiovisuell richtig erkannt in Prozent - Alle	Negative Ränge	46[a]	49,51	2277,50
	Positive Ränge	52[b]	49,49	2573,50

	Bindungen	6[c]		
	Gesamt	104		

a. Alle Videos - Audiovisuell richtig erkannt in Prozent < Alle Videos - Visuell richtig erkannt in Prozent

b. Alle Videos - Audiovisuell richtig erkannt in Prozent > Alle Videos - Visuell richtig erkannt in Prozent

c. Alle Videos - Audiovisuell richtig erkannt in Prozent = Alle Videos - Visuell richtig erkannt in Prozent

Statistik für Test[b]

	Alle Videos - Audiovisuell richtig erkannt in Prozent - Alle Videos - Visuell richtig erkannt in Prozent
Z	-,526[a]
Asymptotische Signifikanz (2-seitig)	,599

a. Basiert auf negativen Rängen.

b. Wilcoxon-Test

Hypothese 4.2: Wer den Ausschnitt audiovisuell erlebt, kann stark negative Emotionen besser unterscheiden als reine Zuhörer und Zuseher.

Hypothese 4.3: Wer den Ausschnitt audiovisuell erlebt, kann stark positive Emotionen besser unterscheiden als reine Zuhörer und Zuseher.

Aufgrund des Sampledesigns war es nicht möglich, Signifikanztests zur Erkennbarkeit einzelner Emotionen anhand von Audio-, Video sowie Audiovideo-Dateien durchzuführen.

Statistiken

		Traurigkeit audiovisuell richtig erkannt in Prozent	Vertrauen audiovisuell richtig erkannt in Prozent	Ekel audiovisuell richtig erkannt in Prozent	Angst audiovisuell richtig erkannt in Prozent
N	Gültig	68	68	68	36
	Fehlend	36	36	36	68

Statistiken

		Wut audiovisuell richtig erkannt in Prozent	Neugierde audiovisuell richtig erkannt in Prozent	Überraschung visuell richtig erkannt in Prozent	Freude audiovisuell richtig erkannt in Prozent
N	Gültig	35	36	71	33
	Fehlend	69	68	33	71

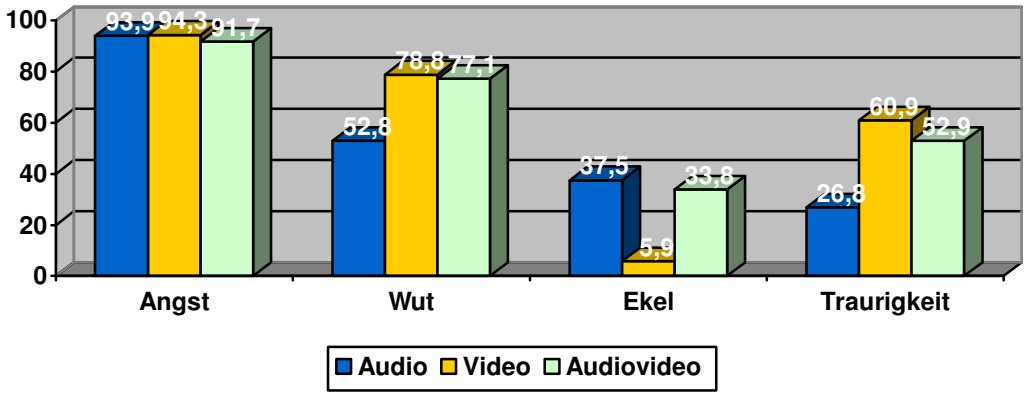
Die Statistiken samt Mittel- und Modalwert lassen dennoch Rückschlüsse darauf zu, ob stark negative sowie stark positive Emotionen durch audiovisuelle Ausschnitte besser zu erkennen sind als durch reine Audio- oder Video-Dateien.

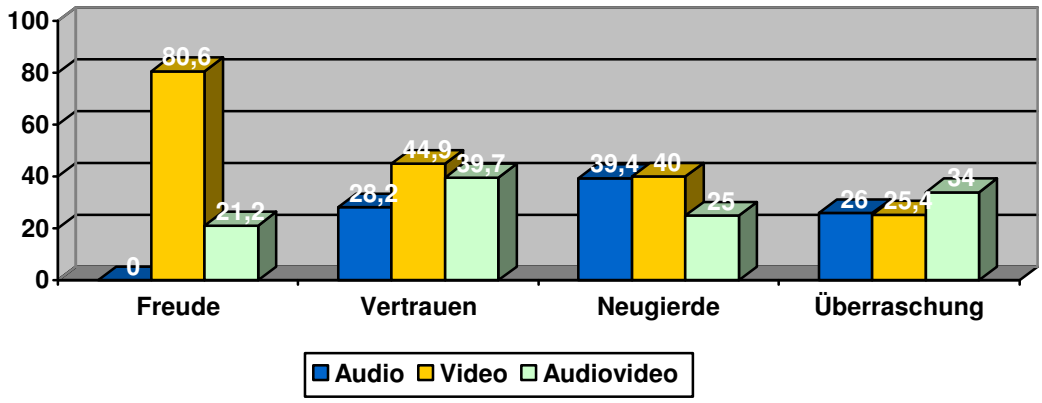
Statistiken

		Ekel audiovisuell richtig erkannt in Prozent	Traurigkeit audiovisuell richtig erkannt in Prozent	Vertrauen audiovisuell richtig erkannt in Prozent	Angst audiovisuell richtig erkannt in Prozent
N	Gültig	68	68	68	36
	Fehlend	36	36	36	68
Mittelwert		33,8235	52,9412	39,7059	91,6667
Modus		,00	100,00	,00	100,00
Standardabweichung		47,66266	50,28453	49,29263	28,03060

Statistiken

		Wut audiovisuell richtig erkannt in Prozent	Neugierde audiovisuell richtig erkannt in Prozent	Freude audiovisuell richtig erkannt in Prozent	Überraschung audiovisuell richtig erkannt in Prozent
N	Gültig	35	36	33	69
	Fehlend	69	68	71	35
Mittelwert		77,1429	25,0000	21,2121	34,0580
Modus		100,00	,00	,00	,00
Standardabweichung		42,60430	43,91550	41,51488	42,43746





Wie aus dem Vergleich mit den Tabellen von Forschungsfrage 3 hervorgeht, wurde die Emotion Traurigkeit am besten erkannt anhand einer Video-Datei ohne Ton (60,9 Prozent), gefolgt von der audiovisuellen Datei mit 52,9 Prozent. Durch die Audio-Datei konnten lediglich 26,8 Prozent der Befragten diese Emotion richtig zuordnen. Ekel wurde anhand eines einzelnen audiovisuellen Ausschnitts von 33,8 Prozent der Befragten identifiziert, bei reinen Audio-Dateien lag der Mittelwert bei 37,5 Prozent. Visuell konnten lediglich 5,9 Prozent der Zuseher Ekel erkennen. Die Identifizierungsquote bei Angst lag beim Video mit Ton bei 91,7 Prozent und damit knapp hinter der reinen Audio-Datei mit 93,9 Prozent richtigen Antworten, sowie der Video-Datei ohne Ton mit 94,3 Prozent. Wut erkannten 77,1 Prozent der Untersuchungsteilnehmer anhand des audiovisuellen Ausschnitts richtig, die Quote lag damit knapp unter jener der Video-Zuseher, die die Emotion zu 78,8 Prozent identifizierten. Anhand der stimmlichen Äußerung allein konnten nur 52,8 Prozent der Teilnehmer Wut erkennen.

Hypothese 4.2 kann aufgrund dieser Ergebnisse als nicht korrekt angesehen werden – stark negative Emotionen werden durch Video-Dateien etwa gleich gut erkannt wie anhand

audiovisueller Ausschnitte. Beide Formen sind bei negativen Emotionen jedoch der stimmlichen Äußerung als Kommunikationsweg überlegen.

39,7 Prozent der Zuseher eines audiovisuellen Ausschnitts konnten Vertrauen identifizieren, damit liegt die Quote knapp unter jener der Video-Zuseher mit 44,9 Prozent korrekten Antworten. Auditiv konnten lediglich 28,2 Prozent der Befragungsteilnehmer diese Emotion zuordnen. Freude hingegen wurde lediglich von 21,2 Prozent der Personen richtig erkannt, die die audiovisuelle Datei bewerteten. Dieser Wert liegt damit deutlich näher an jenem der Audio-Hörer, von denen keiner Freude erkannte, als an dem der Personen, die das Video ohne Ton gesehen hatten – von ihnen konnten 80,6 Prozent Freude identifizieren.

Hypothese 4.3 ist somit ebenfalls widerlegt – positive Emotionen werden eher von reinen Zusehern erkannt als von Zusehern, die ebenfalls Ton hören. Die Erkennungsquote positiver Emotionen bei audiovisuellen Dateien liegt dennoch über jener von stimmlichen Äußerungen. Bei den neutralen Emotionen Neugierde und Überraschung zeigte sich ein anderes Bild: Neugierde wurde auditiv von 39,4 Prozent, visuell nahezu ident von 40 Prozent, audiovisuell allerdings nur von 25 Prozent der Teilnehmer erkannt. Überraschung hingegen wurde anhand der Audiovideo-Datei mit 34 Prozent besser erkannt als anhand des stimmlichen Stimulus (26 Prozent) oder des reinen Bildes (25,4 Prozent).

Aus diesen Ergebnissen lässt sich folgende neue Hypothese generieren:
Hypothese 4.4: Emotionen werden anhand audiovisueller Ausschnitte in ähnlichem Ausmaß erkannt wie anhand visueller Ausschnitte, beide Erlebnisformen sind der auditiven Erkennung überlegen.

Forschungsfrage 5: Welche Unterschiede lassen sich in der auditiven Emotionswahrnehmung in Verbindung mit Persönlichkeitsmerkmalen sowie Empathiefähigkeit ermitteln?
Hypothese 5.0: Es gibt keinen Zusammenhang zwischen Persönlichkeitsmerkmalen und dem Erkennen von Emotionen.
Hypothese 5.0 ist auszuschließen, da Hypothese 5.1 verifiziert wurde.

Hypothese 5.1: Personen mit niedrigen Werten im Bereich Extraversion können Emotionen besser zuordnen als solche mit hohen.

Korrelationen

			Alle Videos - Auditiv richtig erkannt in Prozent	FPI-R Extraversion Statine
Spearman-Rho	Alle Videos - Auditiv richtig erkannt in Prozent	Korrelationskoeffizient	1,000	-,271**
		Sig. (2-seitig)	.	,005
		N	104	104
	FPI-R Extraversion Statine	Korrelationskoeffizient	-,271**	1,000
		Sig. (2-seitig)	,005	.
		N	104	104

**. Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

Da die Daten keine Normalverteilung aufwiesen (siehe Forschungsfrage 2) wurde die Korrelation nach Spearman berechnet. Mit einer Korrelation von r= -0,271 auf einem Signifikanzniveau von p=0,01 zeigte sich ein schwach negativer, hoch signifikanter Zusammenhang zwischen der auditiven Emotionserkennung und dem Extraversionswert (die Statinen Normen wurden nach den Vorgaben des Freiburger Persönlichkeitsinventars berechnet).

Verarbeitete Fälle

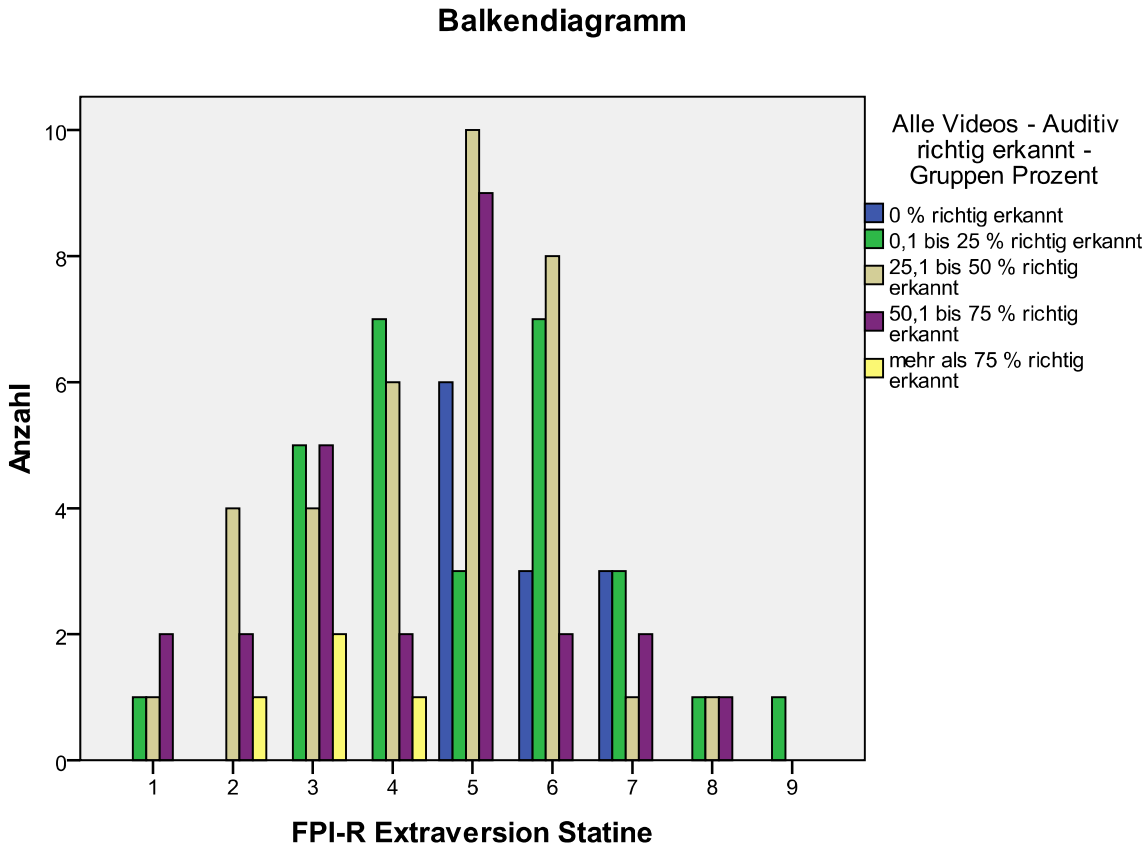
	Fälle					
	Gültig		Fehlend		Gesamt	
	N	Prozent	N	Prozent	N	Prozent
FPI-R Extraversion Statine * Alle Videos - Auditiv richtig erkannt - Gruppen Prozent	104	100,0%	0	,0%	104	100,0%

FPI-R Extraversion Statine * Alle Videos - Auditiv richtig erkannt - Gruppen Prozent Kreuztabelle

Anzahl

		Alle Videos - Auditiv richtig erkannt - Gruppen Prozent					
		0 % richtig erkannt	0,1 bis 25 % richtig erkannt	25,1 bis 50 % richtig erkannt	50,1 bis 75 % richtig erkannt	mehr als 75 % richtig erkannt	Gesamt
FPI-R Extraversion Statine	1	0	1	1	2	0	4
	2	0	0	4	2	1	7
	3	0	5	4	5	2	16

	4	0	7	6	2	1	16
	5	6	3	10	9	0	28
	6	3	7	8	2	0	20
	7	3	3	1	2	0	9
	8	0	1	1	1	0	3
	9	0	1	0	0	0	1
Gesamt		12	28	35	25	4	104



Die Korrelation wird anhand der Kreuztabelle mit zugehörigem Balkendiagramm deutlich: Alle Befragungsteilnehmer, die mehr als 75 Prozent der Emotionen anhand reiner Audio-Dateien identifizierten, wiesen einen Extraversions-Wert von 2 bis 4 auf (auf einer Statinen-Skala von 1 bis 9). Auch diejenigen Fragebogenteilnehmer, die mehr als 50 Prozent der Emotionen anhand einer stimmlichen Äußerung richtig erkannten, liegen im unteren Feld des Extraversions-Werts.

1 steht auf der Skala des Freiburger Persönlichkeitsinventars für Introversion, Zurückhaltung, Überlegtheit und Ernst, 9 für Extraversion, Geselligkeit, Impulsivität und Unternehmungslust.

Der normierte Mittelwert bei den Statinen liegt bei 5 mit einer Standardabweichung 1,96. Bei dieser Untersuchung ergab sich ein Mittelwert von 4,67 mit einer Standardabweichung von 1,692.

Statistiken

FPI-R Extraversion Statine

N	Gültig	104
	Fehlend	0
Mittelwert		4,67
Standardabweichung		1,692

Hypothese 5.1 kann aufgrund der Ergebnisse als bestätigt angesehen werden – Menschen mit niedrigen Extraversionswerten erkennen Emotionen anhand stimmlicher Äußerungen besser als Personen mit hohen Werten.

Der Zusammenhang zwischen Extraversion und der Erkennungsgenauigkeit ist nur bei der auditiven Wahrnehmung gegeben:

Korrelationen

			FPI-R Extraversion Statine	Alle Videos - Visuell richtig erkannt in Prozent	Alle Videos - Audiovisuell richtig erkannt in Prozent
Spearman-Rho	FPI-R Extraversion Statine	Korrelationskoeffizient	1,000	-,013	,164
		Sig. (2-seitig)	.	,900	,095
		N	104	104	104
	Alle Videos - Visuell richtig erkannt in Prozent	Korrelationskoeffizient	-,013	1,000	,177
		Sig. (2-seitig)	,900	.	,072
		N	104	104	104
	Alle Videos - Audiovisuell richtig erkannt in Prozent	Korrelationskoeffizient	,164	,177	1,000
		Sig. (2-seitig)	,095	,072	.
		N	104	104	104

Hypothese 5.2: Personen mit hohen Werten im Bereich Emotionalität können Emotionen besser zuordnen als solche mit niedrigen.

Korrelationen

			Alle Videos - Auditiv richtig erkannt in Prozent	FPI-R Emotionalität Statine
Spearman-Rho	Alle Videos - Auditiv richtig erkannt in Prozent	Korrelationskoeffizient	1,000	,066
		Sig. (2-seitig)	.	,508
		N	104	104
	FPI-R Emotionalität Statine	Korrelationskoeffizient	,066	1,000
		Sig. (2-seitig)	,508	.
		N	104	104

Hypothese 5.2 muss falsifiziert werden, da kein signifikantes Ergebnis (p=0,508) und keine Korrelation (r < = 0,2) existiert.

Hypothese 5.3: Personen mit hohen Werten im Bereich Lebenszufriedenheit können Emotionen besser zuordnen als solche mit niedrigen.

Korrelationen

			Alle Videos - Auditiv richtig erkannt in Prozent	FPI-R Lebenszufriedenheit Statine
Spearman-Rho	Alle Videos - Auditiv richtig erkannt in Prozent	Korrelationskoeffizient	1,000	-,165
		Sig. (2-seitig)	.	,093
		N	104	104
	FPI-R Lebenszufriedenheit Statine	Korrelationskoeffizient	-,165	1,000
		Sig. (2-seitig)	,093	.
		N	104	104

Auch Hypothese 5.3 ist mangels Signifikanz (p=0,093) und Korrelation (r = -0,165) nicht gültig.

Hypothese 5.4: Personen mit hohen Werten im Bereich Soziale Orientierung können Emotionen besser zuordnen als solche mit niedrigen.

Korrelationen

			Alle Videos - Auditiv richtig erkannt in Prozent	FPI-R Soziale Orientierung Statine
Spearman-Rho	Alle Videos - Auditiv richtig erkannt in Prozent	Korrelationskoeffizient	1,000	-,024
		Sig. (2-seitig)	.	,807

		N	104	104
	FPI-R Soziale Orientierung Statine	Korrelationskoeffizient	-,024	1,000
		Sig. (2-seitig)	,807	.
		N	104	104

Hypothese 5.4 ist ebenfalls widerlegt, da die Signifikanz bei p=0,807 lag, die Korrelation bei r = -0,024.

Hypothese 5.5: Personen mit niedrigen Werten im Bereich Gehemmtheit können Emotionen besser zuordnen als solche mit hohen.

Korrelationen

			Alle Videos - Auditiv richtig erkannt in Prozent	FPI-R Gehemmtheit Statine
Spearman-Rho	Alle Videos - Auditiv richtig erkannt in Prozent	Korrelationskoeffizient	1,000	,158
		Sig. (2-seitig)	.	,110
		N	104	104
	FPI-R Gehemmtheit Statine	Korrelationskoeffizient	,158	1,000
		Sig. (2-seitig)	,110	.
		N	104	104

Wegen mangelnder Signifikanz (p=0,110) und Korrelation (r < = 0,2) muss auch Hypothese 5.5 falsifiziert werden.

Hypothese 5.6: Personen mit hohen Werten im Bereich Empathie in realen Situationen können Emotionen besser zuordnen als solche mit niedrigen.

Korrelationen

			Alle Videos - Auditiv richtig erkannt in Prozent	Empathie in realen Situationen - Punktedurchschnitt
Spearman-Rho	Alle Videos - Auditiv richtig erkannt in Prozent	Korrelationskoeffizient	1,000	,086
		Sig. (2-seitig)	.	,383
		N	104	104
	Empathie in realen Situationen - Punktedurchschnitt	Korrelationskoeffizient	,086	1,000
		Sig. (2-seitig)	,383	.
		N	104	104

Hypothese 5.6 ist ebenso ungültig, da das Signifikanzniveau bei p=0,383 lag, die Korrelation bei r < = 0,2.

Hypothese 5.7: Personen mit hohen Werten im Bereich Empathie in fiktiven Situationen können Emotionen besser zuordnen als solche mit niedrigen.

Korrelationen

			Alle Videos - Auditiv richtig erkannt in Prozent	Phantasieempathie - Punktedurchschnitt
Spearman-Rho	Alle Videos - Auditiv richtig erkannt in Prozent	Korrelationskoeffizient	1,000	,012
		Sig. (2-seitig)	.	,908
		N	104	104
	Phantasieempathie - Punktedurchschnitt	Korrelationskoeffizient	,012	1,000
		Sig. (2-seitig)	,908	.
		N	104	104

Hypothese 5.7 muss ebenfalls falsifiziert werden, die sich weder Signifikanz (p=0,908) noch Korrelation (r < = 0,2) ergaben.

Forschungsfrage 6: Welche Unterschiede lassen sich in der auditiven Emotionswahrnehmung in Verbindung mit sozialen Daten ermitteln?

Hypothese 6.0: Es gibt keinen Zusammenhang zwischen sozialen Daten und der Emotionserkennung.

Es ergaben sich für keine der Hypothesen eine Korrelation und ein signifikantes Ergebnis, die eine andere als die Nullhypothese bestätigt hätten. Daher wird für Forschungsfrage 6 lediglich Hypothese 6.0 als verifiziert angesehen, die übrigen Hypothesen sind nicht gültig.

Hypothese 6.1: Frauen können Emotionen besser zuordnen als Männer.

Korrelationen

			Alle Videos - Auditiv richtig erkannt in Prozent	Geschlecht
Spearman-Rho	Alle Videos - Auditiv richtig erkannt in Prozent	Korrelationskoeffizient	1,000	-,004
		Sig. (2-seitig)	.	,969
		N	104	93
	Geschlecht	Korrelationskoeffizient	-,004	1,000

		Sig. (2-seitig)	,969	.
		N	93	93

Diese Hypothese ist falsifiziert, da sich kein signifikantes Ergebnis und keine Korrelation (p=0,969, Korrelation bei r = -0,004) ergaben.

Hypothese 6.2: Ältere Menschen können Emotionen besser zuordnen als jüngere.

Korrelationen

			Alle Videos - Auditiv richtig erkannt in Prozent	Alter
Spearman-Rho	Alle Videos - Auditiv richtig erkannt in Prozent	Korrelationskoeffizient	1,000	-,024
		Sig. (2-seitig)	.	,810
		N	104	103
	Alter	Korrelationskoeffizient	-,024	1,000
		Sig. (2-seitig)	,810	.
		N	103	103

Auch diese Hypothese gilt als falsifiziert, da weder ein signifikantes Ergebnis noch eine Korrelation (p=0,810, r = -0,004) vorlagen.

Hypothese 6.3: Personen mit Geschwistern können Emotionen besser zuordnen als Einzelkinder.

Korrelationen

			Alle Videos - Auditiv richtig erkannt in Prozent	Haben Sie Geschwister?
Spearman-Rho	Alle Videos - Auditiv richtig erkannt in Prozent	Korrelationskoeffizient	1,000	,040
		Sig. (2-seitig)	.	,686
		N	104	103
	Haben Sie Geschwister?	Korrelationskoeffizient	,040	1,000
		Sig. (2-seitig)	,686	.
		N	103	103

Hypothese 6.3 muss ebenfalls falsifiziert werden, da das Signifikanzniveau bei p=0,686 lag sowie die Korrelation bei r = -0,024.

Hypothese 6.4: Personen mit Kindern können Emotionen besser zuordnen als kinderlose Menschen.

Korrelationen

			Alle Videos - Auditiv richtig erkannt in Prozent	Haben Sie Kinder?
Spearman-Rho	Alle Videos - Auditiv richtig erkannt in Prozent	Korrelationskoeffizient	1,000	-,065
		Sig. (2-seitig)	.	,512
		N	104	104
	Haben Sie Kinder?	Korrelationskoeffizient	-,065	1,000
		Sig. (2-seitig)	,512	.
		N	104	104

Auch diese Hypothese ist falsifiziert, da das Signifikanzniveau bei p=0,512 lag, die Korrelation bei Korrelation r = -0.06.

Hypothese 6.5: Eltern von Töchtern können Emotionen besser zuordnen als Eltern von Söhnen.

Korrelationen

			Alle Videos - Auditiv richtig erkannt in Prozent	Haben Sie Kinder? - Gruppen
Spearman-Rho	Alle Videos - Auditiv richtig erkannt in Prozent	Korrelationskoeffizient	1,000	-,131
		Sig. (2-seitig)	.	,184
		N	104	104
	Haben Sie Kinder? - Gruppen	Korrelationskoeffizient	-,131	1,000
		Sig. (2-seitig)	,184	.
		N	104	104

Auch diese Hypothese darf als ungültig angesehen werden (Signifikanz p = 0,13, Korrelation r = -0,131) – auch wenn der Datensatz (wie aus der Kreuztabelle hervorgeht) an Personen, die mindestens eine Tochter bzw. mindestens einen Sohn haben, zu gering ist, um dies zweifelsfrei belegen zu können.

Haben Sie Kinder? - Gruppen * Alle Videos - Auditiv richtig erkannt - Gruppen Prozent Kreuztabelle

Anzahl

	Alle Videos - Auditiv richtig erkannt - Gruppen Prozent	Gesamt

		0 % richtig erkannt	0,1 bis 25 % richtig erkannt	25,1 bis 50 % richtig erkannt	50,1 bis 75 % richtig erkannt	mehr als 75 % richtig erkannt	
Haben Sie Kinder? - Gruppen	Keine Kinder	6	18	19	15	3	61
	Mind. 1 Tochter	1	0	2	5	1	9
	Mind. 1 Sohn	0	2	3	1	0	6
	Mind. 1 Tochter & mind. 1 Sohn	5	8	11	4	0	28
Gesamt		12	28	35	25	4	104

Hypothese 6.6: Personen, die in einer Beziehung leben, können Emotionen besser zuordnen als Singles.

Korrelationen

			Alle Videos - Auditiv richtig erkannt in Prozent	Familienstand - Gruppen
Spearman-Rho	Alle Videos - Auditiv richtig erkannt in Prozent	Korrelationskoeffizient	1,000	,153
		Sig. (2-seitig)	.	,121
		N	104	104
	Familienstand - Gruppen	Korrelationskoeffizient	,153	1,000
		Sig. (2-seitig)	,121	.
		N	104	104

Diese Hypothese muss ebenfalls falsifiziert werden, da das Signifikanzniveau bei p=0,121 lag und die Korrelation bei r < = 0,2.

Forschungsfrage 7: Welchen Einfluss hat der Faktor, ob eine Person eine TV-Serie bereits mindestens einmal gesehen hat, auf die Erkennung der von der Synchronstimme dargestellten Emotion?

Hypothese 7.0: Es besteht kein Unterschied in Bezug auf die Genauigkeit der Emotionserkennung zwischen Menschen, die eine TV-Serie bereits gesehen haben und jenen, die diese Serie noch nicht gesehen haben.

Da keine der beiden Alternativhypothesen eine Korrelation und ein signifikantes Ergebnis aufwiesen, gilt die Nullhypothese als verifiziert.

Statistiken

		Serienkenntnis-Erhebung - Wie oft haben Sie „CSI: Miami“ insgesamt gesehen?	Serienkenntnis-Erhebung - Wie oft haben Sie „Damages“ insgesamt gesehen?	Serienkenntnis-Erhebung - Sehen Sie „CSI: Miami“ regelmäßig im Fernsehen?	Serienkenntnis-Erhebung - Sahen Sie „Damages“ regelmäßig im Fernsehen?
N	Gültig	104	104	99	100
	Fehlend	0	0	5	4
Mittelwert		2,87	1,07	1,60	1,00
Standardabweichung		1,637	,349	1,087	,000

Hypothese 7.1: Zuhörer, die eine TV-Serie bereits gesehen haben, können die Emotion besser zuordnen als Personen, die die Serie noch nie gesehen haben.

Korrelationen

			Haben Sie „CSI: Miami“ schon einmal gesehen?	CSI: Miami - Audio richtig erkannt
Spearman-Rho	Haben Sie „CSI: Miami“ schon einmal gesehen?	Korrelationskoeffizient	1,000	-,168
		Sig. (2-seitig)	.	,160
		N	104	71
	CSI: Miami - Audio richtig erkannt	Korrelationskoeffizient	-,168	1,000
		Sig. (2-seitig)	,160	.
		N	71	71

Korrelationen

			Haben Sie „Damages“ schon einmal gesehen?	Damages - Audio richtig erkannt
Spearman-Rho	Haben Sie „Damages“ schon einmal gesehen?	Korrelationskoeffizient	1,000	,005
		Sig. (2-seitig)	.	,958
		N	104	104
	Damages - Audio richtig erkannt	Korrelationskoeffizient	,005	1,000
		Sig. (2-seitig)	,958	.
		N	104	104

Hypothese 7.1 ist aufgrund eines nicht signifikanten Ergebnisses („CSI: Miami“: p = 0,160, Korrelation r = -0,168; „Damages“: p = 0,958, Korrelation r < = 0,2) nicht gültig.

Hypothese 7.2: Je öfter eine Person die TV-Serie gesehen hat, desto besser kann sie die Emotion zuordnen.

Korrelationen

			CSI: Miami - Audio richtig erkannt	Serienkenntnis-Erhebung - Sehen Sie „CSI: Miami“ regelmäßig im Fernsehen?
Spearman-Rho	CSI: Miami - Audio richtig erkannt	Korrelationskoeffizient	1,000	-,081
		Sig. (2-seitig)	.	,512
		N	71	68
	Serienkenntnis-Erhebung - Sehen Sie „CSI: Miami“ regelmäßig im Fernsehen?	Korrelationskoeffizient	-,081	1,000
		Sig. (2-seitig)	,512	.
		N	68	99

Hypothese 7.2 kann anhand der Auswertung der Serie „CSI: Miami“ – die Signifikanz betrug p = 0,512, Korrelation r = -0,081 – als falsifiziert angesehen werden. Anhand der Serie „Damages“ war keine Korrelation auszuwerten, da keiner der Befragten (4 fehlende Werte, N = 100) die Serie regelmäßig im Fernsehen verfolgt hatte.

Korrelationen

			Serienkenntnis-Erhebung - Sahen Sie „Damages“ regelmäßig im Fernsehen?	Damages - Audio richtig erkannt
Spearman-Rho	Serienkenntnis-Erhebung - Sahen Sie „Damages“ regelmäßig im Fernsehen?	Korrelationskoeffizient	.	.
		Sig. (2-seitig)	.	.
		N	100	100
	Damages - Audio richtig erkannt	Korrelationskoeffizient	.	1,000
		Sig. (2-seitig)	.	.
		N	100	104

Statistiken

	Serienkenntnis-Erhebung - Sahen Sie „Damages“ regelmäßig im Fernsehen?	Damages - Audio richtig erkannt

N	Gültig	100	104
	Fehlend	4	0

Serienkenntnis-Erhebung - Sahen Sie „Damages“ regelmäßig im Fernsehen?

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Verfolg(t)e ich nicht regelmäßig	100	96,2	100,0	100,0
Fehlend	System	4	3,8		
Gesamt		104	100,0		

3.6. Diskussion

Postman stellte 1982 fest:

> „Obwohl man im Fernsehen auch Sprache hört und diese mitunter sogar Wichtigkeit erlangt, ist es gleichwohl das Bild, welches das Bewußtsein [sic] des Zuschauers beherrscht und die entscheidenden Bedeutungen vermittelt. Um es so einfach wie möglich zu sagen: *Die Menschen sitzen als Zuschauer vor dem Fernseher*, nicht als Leser und auch nicht so sehr als Hörer. Sie sehen fern.“[280]

In meiner Untersuchung hat sich diese Annahme Postmans bestätigt – Emotionen werden im Fernsehen offenbar besser mit den Augen als dem Gehör wahrgenommen. Bei nahezu allen Emotionen ist die visuelle Wahrnehmung der auditiven überlegen.

Dennoch zeigte sich, dass Emotionen anhand stimmlicher Äußerungen erkannt werden. Aussagekräftig ist dieses Ergebnis meiner Untersuchung vor allem deshalb, da es sich bei sämtlichen Ausschnitten um nur wenige Sekunden kurze Auszüge handelte und in sechs von sieben Sequenzen nicht nur eine, sondern zwei Emotionen dargestellt wurden. Dieser Höreindruck – die Wahrnehmung und Diskrimination mehrerer Emotionen auf einmal – entspricht meiner Ansicht nach dem natürlichen Hörumfeld, der natürlichen Stimme. Und auch jenem Höreindruck, den wir insbesondere in audiovisuellen Medien gewinnen, da diese stark von fiktionalen Erzählformen geprägt sind, die stark emotionale Inhalte vermitteln. Die Befragten standen demnach nicht nur vor der Wahl, eine bestimmte Emotion zu wählen, sondern hatten die Möglichkeit, eine zweite anzugeben. Ihnen war jedoch weder bekannt, ob bei einer oder mehreren Sequenzen zwei Emotionen gesucht waren, noch um welche

[280] Vgl. Postman (2006), S. 92

Sequenzen es sich dabei handelte. Dass sich bei der Untersuchung trotz dieser Umstände ein Mittelwert von 39,50 Prozent richtiger auditiver Erkennung ergab, zeigt, dass Emotionen über die Stimme trotz möglicherweise widriger Umstände wahrgenommen und erkannt werden. Betrachtet man den Mittelwert unter dem Aspekt, dass bereits eine von zwei wahrgenommenen Emotionen in einer Sequenz als positive Erkennung ausreicht, liegt der Mittelwert gar bei 68,60 Prozent. Dieser Wert liegt sogar über jenem von etwa 60 Prozent, den Scherer[281] bei einem Vergleich von 16 Studien zur vokalen Identifikation von Emotionen errechnete, wo jeweils nur eine von vier bis zehn Emotionen gesucht war.

Ebenfalls eine Erkenntnis dieser Untersuchung ist, dass das auditive Erkennen von Emotionen – zumindest in den Medien – geschlechterunabhängig ist und sich auch keine altersspezifischen Unterschiede zeigen. Eine Einschränkung auf das mediale Erleben von emotionalen, jedoch fiktionalen Inhalten ist nötig, da sich diese Erkenntnis mit jener deckt, dass Männer und Frauen zwar auf interpersonale Erlebnisse emotional unterschiedlich reagieren, auf unpersönliche jedoch ähnlich.[282] Es ist daher möglich, dass Frauen auf vokale Emotionsausdrücke in interpersonaler Kommunikation stärker reagieren oder diese besser deuten können als Männer – wissenschaftlich ist dies allerdings nicht erforscht.
Ob eine Person Kinder hat oder mit Geschwistern aufgewachsen ist, hat auf die Erkennung von Emotionen in der Stimme ebenso wenig Einfluss wie Geschlecht oder Alter. Aus diesen Fakten lässt sich der Schluss ziehen, dass das Erkennen von Gefühlen relativ unabhängig ist von sozialen Daten. Tiefgründige Fragen zur Sozialisationsphase der Befragungsteilnehmer hätten den Rahmen dieser Untersuchung allerdings gesprengt – die Frage, ob das soziale Umfeld eines Kindes, die Eltern oder der Umgang und das Heranführen an die Medien die Emotionserkennung in der Stimme beeinflussen, könnte einen Ansatz für weiterführende Forschung bieten.

Die spezifische Dekodierungskompetenz lässt sich laut meinen Erkenntnissen nicht von der interpersonalen Ebene auf die mediale Ebene übertragen. Wie oft jemand eine Fernsehserie bereits gesehen hat, ob er also mit den Stimmen und ihrer speziellen Ausdrucksform vertraut ist oder nicht, hat keinen Einfluss auf die Erkennungsgenauigkeit von Emotionen. Dieses Ergebnis deckt sich mit dem Konsens der Stimmforschung, dass grundlegende stimmliche Merkmale von Emotionen unabhängig von individuellen Stimmmerkmalen wahrgenommen

[281] Vgl. Scherer, in: Scherer (1982), S. 303
[282] Vgl. Merten (2003), S. 153

werden – da aufgrund der Unterschiedlichkeit von Stimmen sonst keine Dekodierung und somit Kommunikation möglich wäre.

Anhand des Freiburger Persönlichkeitsinventars (FPI-R) hat sich lediglich ein Persönlichkeitsmerkmal als mitentscheidend für die Emotionserkennung herauskristallisiert: die Extraversion. Personen mit hohen Werten, die laut Test gerne ausgehen und schnell Freunde finden, lebhaft und gesprächig sind, hatten mehr Schwierigkeiten bei der Emotionserkennung anhand der Stimme als Personen, die laut FPI-R zurückhaltend und nicht sehr gesellig, wenig mitteilsam und ernst sind. Eine Interpretationsmöglichkeit dieses Ergebnisses ist, dass wer viel und gerne spricht und sich mit Menschen umgibt, deshalb kein guter Menschenkenner sein muss. Allerdings lässt das Resultat auch eine andere Auslegung zu: Introvertierte Personen, die sich nicht gerne mit Menschen umgeben, verbringen möglicherweise mehr Zeit alleine vor dem Fernseher, haben so intensiveren Zugang zur medialen Vermittlung von Emotion und können daher Emotionen in TV-Serien anhand der Stimme besser identifizieren. Gegen diese These spricht jedoch, dass der Extraversions-Wert lediglich mit den Ergebnissen der auditiven Stimuli korreliert, nicht aber den rein visuellen oder audiovisuellen.

Bei den einzelnen Emotionen zeigte sich ein differenziertes Bild. So wurde zum Beispiel Freude in dieser Untersuchung auditiv von keinem Befragten erkannt – was einerseits durch die Auswahl der Sequenz bedingt sein könnte, die eher unterschwellige Freude bzw. Schadenfreude und somit nicht die reine, positive Emotion darstellen könnte, die typischerweise mit dem Emotionswort „Freude“ assoziiert wird. Allerdings stellte bereits Scherer[283] fest, dass negative Emotionen anhand der Stimme besser erkannt werden als positive. Interessanterweise wurde Freude im Gegensatz dazu von 80,6 Prozent der reinen Zuseher identifiziert, allerdings nur von 21,2 Prozent derjenigen mit audiovisuellem Stimulus. Dieses Ergebnis lässt vermuten, dass Stimme bei dieser Emotion im normalen Alltag sowie beim Konsum audiovisueller Medien eine größere Rolle spielt als die visuelle Wahrnehmung. Angst wurde von 93,9 Prozent der Befragten anhand der Stimme erkannt, was die These Scherers unterstützt. Ekel erreichte einen Mittelwert von 37,5 Prozent, im Gegensatz dazu erkannten lediglich 5,9 Prozent der rein visuellen Zuseher diese Emotion. Hier lässt sich ebenfalls, wie bei der Freude, ein Zusammenhang zwischen der Emotionserkennung und der Stimme herstellen –die Erkennungsquote von Ekel lag anhand der audiovisuellen Vorführung

[283] Vgl. Scherer, in: Scherer (1982), S. 303

derselben Sequenz bei 33,8 Prozent, was darauf schließen lässt, dass die Stimme auch bei der Erkennung von Ekel eine entscheidende Rolle spielt.
Im Vergleich mit den Ergebnissen von Williams und Stevenson[284] aus dem Jahr 1982 zeigen sich sowohl Übereinstimmungen als auch Diskrepanzen. So wurde Traurigkeit in der Untersuchung 1982 zu 73 Prozent erkannt, bei dieser jedoch nur zu 26,8 Prozent. Ärger wurde nahezu gleich gut erkannt – 52,8 Prozent bei dieser zu 51 Prozent bei der Studie von Williams und Stevenson. Angst wurde von meinen Befragten zu 93,9 Prozent identifiziert, 1982 jedoch nur von 27 Prozent. Freude erreicht bei Williams und Stevenson mit nur 28 Prozent ebenfalls einen schlechten Erkennungswert, in meiner Befragung wurde diese Emotion gar nicht erkannt.

Grundsätzlich zeigte der Vergleich zwischen den Darbietungsformen, dass Emotionen sowohl durch visuelle als auch audiovisuelle Stimuli besser erkannt werden als anhand der Stimme, allerdings brachte die Zusammenführung der Informationen im Vergleich zu den reinen Video-Dateien keinen höheren Identifizierungsgrad der Emotion mit sich. Dieses Ergebnis steht im Widerspruch zur Forschung von Kreifelts et al.[285] (siehe 3.1), wo audiovisuelle Stimuli mit 86 Prozent signifikant besser erkannt wurden als rein visuelle mit 75 Prozent. Der Abstand zu den korrekt identifizierten Emotionen anhand der Stimme, 56 Prozent, scheint hingegen größer als bei meiner Untersuchung, wo Audio- durchschnittlich zu 39,50 Prozent, Video- zu 48,50 Prozent und audiovisuelle Dateien zu 49,92 Prozent Erkennungsquote führten. Allerdings sind die Ergebnisse aufgrund des unterschiedlichen Untersuchungsdesigns nicht direkt vergleichbar, so wurden bei der Untersuchung von Kreifelts et al. insgesamt 168 (extra produzierte) Stimuli eingesetzt und es standen lediglich sechs Emotionen plus die Option „neutral“ zur Auswahl.

Der Unterschied in der Erkennungsquote könnte allerdings zum Teil auch dadurch erklärt werden, dass bei meiner Untersuchung der Fokus auf der medialen Vermittlung von Emotion lag, was die Befragten wussten. Möglicherweise lag der Fokus der Zuseher – besonders bei audiovisuellen Dateien – mehr auf der Erkennung der Sequenz (bei CSI: Miami), der handelnden Personen, kurz: den medienspezifischen Inhalten. Möglicherweise ist es einfacher, sich den Gewohnheiten des Mediums Fernsehen – dem Verlangen, die Geschichte zu verstehen, Lücken zu füllen – zu entziehen und sich lediglich auf die Emotionserkennung

[284] Vgl. Williams / Stevens, in Scherer (1982), S. 313 ff.
[285] Vgl. Kreifelts in: NeuroImage (2007), S. 1445 ff.

zu konzentrieren, wenn reine Audio- oder Video-Datei vorliegen statt der audiovisuellen, die unwillkürlich mit dem Ritual Fernsehen in Verbindung gebracht wird.

Einen ähnlichen Ansatz verfolgt Schlimbach[286], sie geht davon aus, dass die Verarbeitung emotionaler Information leichter fällt, je weniger Kanäle diese übermitteln, da ansonsten viel Energie in die Selektion fließt.

> „Bei audiovisuellen Medien muss dagegen mehr Energie aufgewendet werden, um relevante externe Informationen zu identifizieren und zu selektieren. Hierbei steht den internen Systemprozessen weniger Energie zur Verfügung. (…) Der emotionale Informationsverarbeitungsmodus wird sich deshalb vermutlich umso weniger negativ auf den Umfang von Selektionsprozessen auswirken, je weniger Informationskanäle ein Medium aufweist.“[287]

Sie stellt jedoch fest, dass eine Video- grundsätzlich mehr Information als eine Audio-Datei bietet, da letztere lediglich einen Informationsstrom liefert, ein Video jedoch z.B. einen Ort und eine Person präsentiert. Kombiniert man die Erkenntnisse zur Überforderung durch mehrere Kanäle sowie dass visuell mehr Informationen angeboten und verarbeitet werden können als auditiv, so decken sich diese mit meinen Forschungsergebnissen. Interessant wäre in diesem Zusammenhang aufzuklären, ob die Kombination zweier Kanäle im Medium zu einer verstärkten Aktivierung des Zusehers führt. So konnten etwa Kreifelts et al.[288] nachweisen, dass wütende Stimmen eine Gehirnregionen mit dem Sitz eines Teils des Gedächtnisses verstärkt anregen. Möglicherweise führt diese Stimulation zum unbewussten Aufrufen von Gedächtnisinhalten, was bei der reinen Verarbeitung von auditiven Stimuli die Emotionserkennung nicht beeinflusst – soll jedoch zusätzlich zur auditiven noch die zugehörige visuelle Information verarbeitet werden, könnte dies eine Verminderung der Aufmerksamkeit und der Erkennungsgenauigkeit von Emotionen nach sich ziehen.

Emotionen im Fernsehen entsprechen nicht jener Ausdrucksform, die den interpersonellen Umgang prägt. Und dennoch nimmt das Fernsehen gerade auch durch die einfache Vermittlung von Emotionen in selbst bestimmbarer Dosis einen nicht unwesentlichen Teil des modernen Lebens ein. Es bleibt die Frage, ob der Mensch sich im Laufe der Sozialisation und der Adaptierung an das Medium auch an die Ausdrucksformen von Emotionalität gewöhnt. Bei der Bildsprache ist dies gängige Praxis und wird von den meisten Menschen nicht aktiv

[286] Vgl. Schlimbach (2007), S. 195 f.
[287] Vgl. Ebenda, S. 195 f.
[288] Vgl. Kreifelts in: NeuroImage (2007), S. 1445 ff.

wahrgenommen. Zum Beispiel, dass die Froschperspektive eingesetzt wird, um Dominanz und Einschüchterung auszudrücken – und vom Zuseher auch so verstanden wird. Fraglich ist also, ob sich Stimme dem Medium anpasst (wie Sänger auf einer großen Bühne) oder unabhängig bleibt. Ob Emotionen – vom schauspielerischen Element generell abgesehen – ähnlich dem natürlichen Ausdruck vermittelt oder möglicherweise sogar medienspezifisch wahrgenommen werden. Interessant wäre in diesem Zusammenhang, wenn emotionale Sequenzen aus dem Fernsehen ausgewählt und ebenjene Texte mit denselben emotionalen Vorgaben von Theaterschauspielern sowie Laien aufgenommen würden. Möglicherweise würde eine vergleichende Untersuchung mit den stimmlichen Ausdrücken dieser Gruppen helfen aufzuklären, ob Emotionserkennung anhand der Stimme unabhängig vom Medium passiert – sowohl auf Enkodierungs- als auch auf Dekodierungsseite.

4. Literatur:

Battacchi, Marco Walter / Suslow, Thomas / Renna, Margherita (1997): *Emotion und Sprache. Zur Definition der Emotion und ihren Beziehungen zu kognitiven Prozessen, dem Gedächtnis und der Sprache*. 2., durchges. Auflage. Frankfurt am Main / Wien (u.a.): Lang.

Bignell, Jonathan (2008): *An Introduction to Television Studies*. 2. Auflage. London (u.a.): Routledge.

Birbaumer, Niels / Schmidt, Robert F.: *Kognitive Funktionen und Denken*. In: Schmidt, Robert F. / Schaible, Hans-Georg (Hg.) (2006): *Neuro- und Sinnesphysiologie*. 5., neu bearbeitete Auflage. Heidelberg: Springer Medizin Verlag.

Bräutigam, Thomas (2001): *Lexikon der Film- und Fernsehsynchronisation. Stars und Stimmen: wer synchronisiert wen in welchem Film? Mehr als 2000 Filme und Serien mit ihren deutschen Synchronsprechern*. Berlin: Lexikon Imprint Verlag.

Bruun Vaage, Margrethe: *Empathie. Zur episodischen Struktur der Teilhabe am Spielfilm.* In: Schick, Thomas / Ebbrecht, Tobias (Hg.) (2008): *Emotion – Empathie – Figur: Spielformen der Wahrnehmung.* Berlin: Vistas.

Burton, Graeme (2000): *Talking Television. An introduction to the study of television*. London: Arnold.

Chamberlain, David (1990): *Woran Babys sich erinnern. Die Anfänge unseres Bewußtseins im Mutterleib*. München: Kosel.

Chion, Michel (1994): *Audio-vision. Sound on screen*. Vom Französischen ins Englische übersetzt von Claudia Gorbman. New York: Columbia University Press.

Chion, Michel (1999): *The voice in cinema*. Vom Französischen ins Englische übersetzt von Claudia Gorbman. New York: Columbia University Press.

Clark, Herbert H. / Clark, Eve V. (1977): *Psychology and language. An introduction to psycholinguistics*. New York (u.a.): Harcourt Brace Jovanovich.

Denk, Doris-Maria / Brunner, Elke / Bigenzahn, Wolfgang: *Grundlagen III: Entwicklung von Sprache und Sprechen*. In: Friedrich, Gerhard / Bigenzahn, Wolfgang / Zorowka, Patrick (2000): *Phoniatrie und Pädaudiologie. Einführung in die medizinischen, psychologischen und linguistischen Grundlagen von Stimme, Sprache und Gehör*. 2., vollständig überarb. und erweiterte Auflage. Bern (u.a.): Huber.

Diendorfer-Radner, Gabriele: *Grundlagen IV: Entwicklungs- und Wahrnehmungspsychologie*. In: Friedrich, Gerhard / Bigenzahn, Wolfgang / Zorowka, Patrick (2000): *Phoniatrie und*

Pädaudiologie. Einführung in die medizinischen, psychologischen und linguistischen Grundlagen von Stimme, Sprache und Gehör. 2., vollständig überarb. und erweiterte Auflage. Bern (u.a.): Huber.

Eckert, Hartwig / Laver, John (1994): *Menschen und ihre Stimmen. Aspekte der vokalen Kommunikation*. Weinheim: Beltz Psychologie Verlags Union.

Fahrenberg, Jochen / Hampel, Rainer / Selg, Herbert (2001): *Das Freiburger Persönlichkeitsinventar. FPI-R.* 7., überarbeitete und neu normierte Auflage. Göttingen (u.a.): Hogrefe.

Faulstich, Werner (2005): *Filmgeschichte*. Paderborn: Wilhelm Fink.

Faulstich, Werner (2008): *Grundkurs Fernsehanalyse*. Paderborn: Wilhelm Fink.

Friedrich, Gerhard: *Grundlagen I: Physiologie von Stimme und Sprechen.* In: Friedrich, Gerhard / Bigenzahn, Wolfgang / Zorowka, Patrick (2000): *Phoniatrie und Pädaudiologie. Einführung in die medizinischen, psychologischen und linguistischen Grundlagen von Stimme, Sprache und Gehör*. 2., vollständig überarb. und erweiterte Auflage. Bern (u.a.): Huber.

Geißner, Hellmut (Hg.) (2002): *Das Phänomen Stimme in Kunst, Wissenschaft, Wirtschaft*. St. Ingbert: Röhrig Universitätsverlag.

Geißner, Hellmut (1981): *Sprechwissenschaft. Theorie der mündlichen Kommunikation*. Kronberg/Ts.: Scriptor Verlag.

Geißner, Hellmut (1991): *Vor Lautsprecher und Mattscheibe. Medienkritische Arbeiten 1965 – 1990*. St. Ingbert: Röhrig Universitätsverlag

Goldstein, Bruce E. / Irtel, Hans (Hg.) (2008): *Wahrnehmungspsychologie. Der Grundkurs*. 7. Auflage. Berlin / Heidelberg: Spektrum Akademischer Verlag.

Gregory, Richard L. (2001): *Auge und Gehirn. Psychologie des Sehens*. 5. Auflage. Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag.

Gruhn, Wilfried (2005): *Der Musikverstand. Neurobiologische Grundlagen des musikalischen Denkens, Hörens und Lernens*. 2., neu überarbeitete Auflage. Hildesheim (u.a.): Georg Olms Verlag.

Gundermann, Horst (1994): *Phänomen Stimme*. München (u.a.): Reinhardt.

Grau, Alexander: *Zeitpunkte, Zeitfenster, Zeiträume – Wie das Gehirn unsere Wahrnehmung organisiert*. In: Felsmann, Klaus-Dieter (Hg.) (2008): *Der Rezipient im Spannungsfeld von Zeit und Medien. Erweiterte Dokumentation zu den 11. Buckower Mediengesprächen 2007*. München: koepaed.

Habermann, Günther (2001): *Stimme und Sprache. Eine Einführung in ihre Physiologie und Hygiene ; für Ärzte, Sänger, Pädagogen und alle Sprechberufe.* 3., unveränd. Auflage. Stuttgart (u.a.): Thieme.

Hickethier, Knut: *Die kulturelle Bedeutung medialer Emotionserzeugung.* In: Bartsch, Anne / Eder, Jens / Fahlenbrach, Kathrin (Hg.) (2007): *Audiovisuelle Emotionen. Emotionsdarstellung durch audiovisuelle Medienangebote.* Köln: Herbert von Halem Verlag.

Hickethier, Knut (1996): *Film- und Fernsehanalyse.* 2., überarb. Auflage. Stuttgart / Weimar: Metzler.

Hickethier, Knut (Hg.) (2005): *Kriminalfilm*. (= Reihe Filmgenres). Stuttgart: Philipp Reclam jun.

Hülshoff, Thomas (1999): *Emotionen. Eine Einführung für beratende, therapeutische, pädagogische und soziale Berufe*. München (u.a.): Reinhardt.

Johnson-Laird, Philip (1996): *Der Computer im Kopf. Formen und Verfahren der Erkenntnis*. Ins Deutsche übersetzt von Friedrich Griese. Deutsche Erstausgabe. München: Deutscher Taschenbuch Verlag.

Krah, Hans: *Kommunikationssituation, Sprechsituation, Semantik.* In: Krah, Hans / Titzmann, Michael (Hg.) (2006): *Medien und Kommunikation. Eine interdisziplinäre Einführung*. Passau: Verlag Karl Stutz.

Mangold, Roland: *Anforderung an die Erklärung audiovisueller Emotionen und Vorstellung eines integrativen Ansatzes*. In: Bartsch, Anne / Eder, Jens / Fahlenbrach, Kathrin (Hg.) (2007): *Audiovisuelle Emotionen. Emotionsdarstellung durch audiovisuelle Medienangebote.* Köln: Herbert von Halem Verlag.

Mathelitsch, Leopold / Friedrich, Gerhard (1995): *Die Stimme. Instrument für Sprache, Gesang und Gefühl*. Berlin / Heidelberg: Springer.

Merten, Jörg (2003): *Einführung in die Emotionspsychologie*. Stuttgart: Kohlhammer.

Mikunda, Christian (2002): *Kino spüren: Strategien emotionaler Filmgestaltung*. Wien: WUV Universitätsverlag.

Monaco, James (2000): *Film verstehen. Kunst, Technik, Sprache, Geschichte und Theorie des Films und der Medien. Mit einer Einführung in Multimedia.* 2. Auflage. Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag.

Ohler, Peter (1994): *Kognitive Filmpsychologie. Verarbeitung durch mentale Repräsentation narrativer Filme*. Münster: MAkS Publikation.

Papoušek, Hamuš / Papoušek, Mechthild: *Zur Frühentwicklung der Kommunikation*. In: Scherer, Klaus R. (Hg.) (1982): *Vokale Kommunikation. Nonverbale Aspekte des Sprachverhaltens*. Weinheim (u.a.): Beltz.

Plutchik, Robert (2003*)*: *Emotions and Life. Perspectives from psychology, biology, and evolution*. Washington, DC (u.a.): American Psychological Association.

Postman, Neil (2006): *Das Verschwinden der Kindheit*. Übersetzt von Reinhard Kaiser. 16. Auflage. Frankfurt am Main: Fischer Taschenbuch Verlag.

Rest, Franz / Amanshauser, Wolfgang: *Die Explosion der Bilder. Entwicklung der Programmstrukturen im österreichischen Fernsehen.* In: Fabris, Hans Heinz / Luger, Kurt (Hg) (1988): *Medienkultur in Österreich. Film, Fotografie, Fernsehen und Video in der Zweiten Republik*. Wien / Köln / Graz: Böhlau Verlag.

Römer, Claus (1994): *Schall und Raum. Eine kleine Einführung in die Welt der Akustik*. Berlin (u.a.): VDE Verlag.

Scherer, Klaus R.: *Die vokale Kommunikation emotionaler Erregung*. In: Scherer, Klaus R. (Hg.) (1982): *Vokale Kommunikation. Nonverbale Aspekte des Sprachverhaltens*. Weinheim (u.a.): Beltz.

Schlimbach, Inga (2007): *Emotionen und Informationsverarbeitung bei der Medienrezeption. Entwicklung und Überprüfung eines neuen Ansatzes.* München: Verlag Reinhard Fischer.

Siegmüller, Julia / Bartels, Henrik (Hg.) (2006): *Leitfaden Sprache Sprechen Stimme Schlucken*. München / Jena: Urban & Fischer Verlag.

Smith, Murray: *Was macht es für einen Unterschied? Wissenschaft, Gefühl und Film.* In: Bartsch, Anne / Eder, Jens / Fahlenbrach, Kathrin (Hg.) (2007): *Audiovisuelle Emotionen. Emotionsdarstellung durch audiovisuelle Medienangebote.* Köln: Herbert von Halem Verlag.

Sonnenschein, David (2001): Sound Design. *The expressive power of music, voice, and sound effects in cinema*. Studio City: Michael Wiese Productions.

Stam, Robert / Burgoyne, Robert / Flitterman-Lewis, Sandy (1994): *New vocabularies in film semiotics. Structuralism, poststructuralism and beyond*. Nachdruck. London (u.a.): Routledge.

Stassen, Hans H. (1995): *Affekt und Sprache. Stimm- und Sprachanalysen bei Gesunden, depressiven und schizophrenen Patienten*. Berlin (u.a.): Springer.

Steininger, Christian / Woelke, Jens (Hg.) (2008): *Fernsehen in Österreich 2008.* Konstanz: UVK Verlagsgesellschaft.

Stelzig, Gertraud: *Zum Phänomen Stimme. Vortrag.* In: Lotzmann, Geert (Hg.) (1997): *Die Sprechstimme. Entstehung – Bildung – Gestaltung – Vorbeugung – Untersuchung – Behandlung*. 1. Auflage. Ulm (u.a.): Fischer.

Tischer, Bernd (1993): *Die vokale Kommunikation von Gefühlen*. (= *Fortschritte der psychologischen Forschung*, Band 18). Weinheim: Beltz Psychologie Verlags Union.

Ulich, Dieter: *Ein persönlichkeitspsychologisches Modell der Entstehung von Gefühlszuständen*. In: Ulich, Dieter / Mayring, Philipp (2003): *Psychologie der Emotionen.* 2., überarbeitete und erweiterte Auflage. Stuttgart: Verlag W. Kohlhammer.

Westphal, Kristin (2002): *Wirklichkeiten von Stimmen. Grundlegung einer Theorie der medialen Erfahrung*. Frankfurt am Main / Wien (u.a.): Lang.

Wickel, Hans Hermann / Hartogh, Theo (2006): *Musik und Hörschäden. Grundlagen für Prävention und Intervention in sozialen Berufsfeldern*. Weinheim / München: Juventa Verlag.

Williams, Carl E. / Stevens, Kenneth N.: *Akustische Korrelate diskreter Emotionen*. In: Scherer, Klaus R. (Hg.) (1982): *Vokale Kommunikation. Nonverbale Aspekte des Sprachverhaltens*. Weinheim (u.a.): Beltz.

Zimmer, Katharina (1998): *Erste Gefühle. Das frühe Band zwischen Kind und Eltern*. München: Kosel.

Zenner, Hans-Peter: *Hören*. In: Schmidt, Robert F. / Schaible, Hans-Georg (Hg.) (2006): *Neuro- und Sinnesphysiologie*. 5., neu bearbeitete Auflage. Heidelberg: Springer Medizin Verlag.

Zettl, Herbert (1995): *Sight, sound, motion. Applied media aesthetics*. 2. Auflage. Belmont: Wadsworth Publishing Company.

Zorowka, Patrick / Höfler, Heribert: *Grundlagen V: Ohr und Gehör.* In: Friedrich, Gerhard / Bigenzahn, Wolfgang / Zorowka, Patrick (2000): *Phoniatrie und Pädaudiologie. Einführung in die medizinischen, psychologischen und linguistischen Grundlagen von Stimme, Sprache und Gehör*. 2., vollständig überarb. und erweiterte Auflage. Bern (u.a.): Huber.

Zeitschriften:

Leibetseder, Max et al.: *E-SKala: Fragebogen zur Erfassung von Empathie – Beschreibung und psychometrische Eigenschaften*. In: *Zeitschrift für Differentielle und Diagnostische Psychologie* (2001), 22. Jahrgang, Heft 4.

Medien in Österreich (2006), Wien, 2., überarb. Auflage – Bundespressedienst Österreich – Bundeskanzleramt, Bundespressedienst, Autor: Mag. Andreas Ulrich, Wien

Ethofer, Thomas et al.: *Effects of prosodic emotional intensity on activation of associative auditory cortex.* In: *NeuroReport* (2006), 17. Jahrgang, Heft 3. URL: http://labnic.unige.ch/nic/papers/Ethofer_NeuroReport2006.pdf Stand: 20.3. 2009.

Kreifelts, Benjamin et al.: *Audiovisual integration of emotional signals in voice and face: An event-related fMRI study*. In: *NeuroImage* (2007), Heft 37. URL: http://labnic.unige.ch/nic/papers/Kreifelts_NeuroImage2007.pdf Stand: 10. 5. 2009.

Wildgruber, Dirk et al.: *Identification of emotional intonation evaluated by fMRI*. In: *NeuroImage* (2005), Heft 24. URL: http://labnic.unige.ch/nic/papers/Wildgruber_NeuroImage2005.pdf Stand: 21. 5. 2009.

Online-Dokumente:

BBC Online (2006): *CSI show ‚most popular in the world'*. URL: http://news.bbc.co.uk/1/hi/entertainment/5231334.stm Stand: 18. 3. 2009.

Media-Analyse (2008). URL: http://www.media-analyse.at/ Stand: 7. 5. 2009.

Medienforschung ORF. URL: http://mediaresearch.orf.at/index2.htm?fernsehen/fernsehen_nutzungsverhalten.htm

Tabellen und Diagramme:

Sämtliche verwendeten Tabellen und Diagramme wurden von mir persönlich anhand der aus meiner Untersuchung gewonnenen Daten entweder per SPSS oder in Microsoft Wort erstellt.

5. Anhang

Im Anhang finden Sie:

Abstract
Einstellungsanalysen der ausgewählten Stimuli
Online-Fragebogen
Lebenslauf

Abstract

Die Diplomarbeit dreht sich um die Frage, wie Stimme als Medium im Medium wirkt, wie sie wahrgenommen wird und welche Emotionen durch sie vermittelt und erkannt werden können. Die theoretischen Grundlagen bildeten die Überlegungen zum Vokozentrismus, dass Stimme einen hohen Stellenwert in der auditiven Wahrnehmung einnimmt. Der große Einfluss von Sound in Film und Fernsehen hat der Stimme Aufschwung beschert. Wie Emotionen anhand der Stimme en- und dekodiert werden, wurde ebenso beleuchtet wie Empathie und Involvierung beim Fernsehen.

Das Forschungsinteresse des empirischen Teils drehte sich um die Frage, ob Zuhörer anhand sehr kurzer stimmlicher Stimuli aus (deutsch synchronisierten) Fernsehserien erkennen können, welche Emotionen vermittelt werden sollten. Außerdem wurde beleuchtet, welche Emotionen anhand der Stimme klar erkannt und welche schwer zu identifizieren sind. Diese Ergebnisse wurden jenen von rein visuellen sowie audiovisuellen Stimuli gegenübergestellt. Es wurde eine quantitative Erhebung mittels Online-Fragebogen durchgeführt, 104 gültige Fragebögen wurden ausgewertet.

Die Ergebnisse der Untersuchung zeigen, dass Zuhörer Emotionen anhand der Stimme erkennen können, allerdings signifikant schlechter als reine Zuseher und audiovisuelle Zuschauer. Stark negative Stimuli werden anhand der Stimme besonders gut identifiziert, positive hingegen schlecht. Stimme scheint in Bezug auf die audiovisuelle Wahrnehmung von Emotionen besonders bei Ekel (verstärkend) und Freude (abschwächend) große Wirkung zu haben. Als einziges Persönlichkeitsmerkmal der Untersuchung korrelierten niedrige Extraversions-Werte mit hohen Werten bei der auditiven Emotionswahrnehmung. Das Ergebnis könnte sowohl auf eine höhere Dekodierungskompetenz introvertierter Personen deuten als auch auf verstärkten Medienkonsum und dadurch bessere Emotionserkennung bei Synchronsprechern.

The diploma thesis centres around the question, how voice takes effect as medium in the medium, how it is perceived and which emotions can be communicated and distinguished through it. The theoretical foundations consisted of the thought of vococentrism, that voice is of high value in the auditory perception. The importance of sound in film and television has

brought an upturn for the voice. How emotions are en- and decoded was examined, as well as empathy and involvement whilst watching TV.

The research interest focused on the question, whether or not listeners can identify emotions on the basis of very short vocal stimuli taken from (synchronized in German) television series. Furthermore it was examined, which emotions are easily recognized through voice and which are hard to discover. The survey was taken via online-questionnaire, 104 valid forms have been analyzed.

The evaluation shows that listeners are able to identify emotions by means of the voice, however significantly worse than mere viewers as well as audiovisual audience. Highly negative stimuli are distinguished particularly well on the basis of voice, positive stimuli on the other hand are hardly detected. Voice seems to have great effect on the audiovisual perception of emotions when it comes to disgust and happiness. As the only personality trait, low values of extraversion correlated with high values of emotion detection. This result could indicate that introverted persons possess a higher competence in decoding emotions – or that they turn more frequently towards the media and can thus identify emotions expressed by voice actors easier.

Einstellungsanalysen der ausgewählten Stimuli

Damages – „Ich bin so froh, dass du da bist.“ (Traurigkeit, Vertrauen)

Beginn: 15:25 – Ende: 16:22

Nr.	Zeit	Kameraperspektive	Einstellungsgröße	Inhalt	Dialog
1	15:25	Normalsicht	Halbtotale	Kamera fährt nach links, Ellen Parsons Mutter steht in einem Vorraum an einer geschlossenen Tür und erzählt:	„Also du weißt doch, dass es jetzt sonntags immer eine zweite Messe gibt.“
				Man hört Ellens Stimme:	„Dad hat’s mir erzählt.“
				Ihre Mutter spricht weiter:	„Ja, anscheinend sollen dadurch jüngere Leute angezogen werden. Und statt dem traditionellen Chor wollen sie da jetzt Rock’n’Roll-Musik spielen.“
2	15:36	Normalsicht	Halbnah	Ellens Mutter lacht, nippt an einem Glas Wein, die Kamera fährt langsam auf sie zu während sie weiterspricht:	„Der Pastor hat deine Schwester gebeten, eine Band zusammenzustellen. Sie wird endlich ein Rockstar, kannst du dir das vorstellen?“
			Nahe	Die Kamera bleibt vor Ellens Mutter stehen, die	„Ellen?“

				erst wartet und schweigt, schließlich fragt:	
				Als keine Reaktion zu hören ist, klopft Ellens Mutter an die Zimmertür und öffnet sie, die Kamera fährt mit in den Raum	
3	15:50	Leichte Draufsicht	Halbtotale	Ellen sitzt mit ihrem Brautkleid auf dem Bett und weint. Ihre Mutter geht auf sie zu und fragt:	„Elli, was ist denn los?
				Ellens Mutter stellt ihr Weinglas auf einer Kommode ab und setzt sich neben Ellen auf das Bett, sie legt den Arm um sie und fragt:	„Oh… Willst du mein Kleid nicht anziehen?“
4	16:02	Normalsicht	Nahe	Ellen lächelt gequält, weint aber weiter, kuschelt sich an ihre Mutter und sagt:	**„Ich bin so froh, dass du da bist!“**
				Die Mutter nimmt sie in den Arm und beruhigt sie:	„Oooh, das bin ich auch, Liebes. Das bin ich doch auch.“
		Leichte Draufsicht	Nahe	Ellen weint weiter in den Armen ihrer Mutter, die Kamera fährt näher an ihr Gesicht	

Damages – „Nein, du hast gegrunzt wie ein Höhlenmensch!“ (Wut, Abscheu/Ekel)

Beginn: 20:30 – Ende: 21:28

Nr.	Zeit	Kameraperspektive	Einstellungsgröße	Inhalt	Dialog

1	20:30	Normalsicht	Großaufnahme	Ein Schlafzimmer – die Kamera schwenkt nach links auf Davids Gesicht, der voller Verachtung ist: („sie“ ist Ellens Chefin Patty, Katie ist Davids Schwester. Katie wurde von Patty für ihre Zwecke missbraucht und dann fallengelassen, was David nicht verzeihen kann.)	„Und dann fängt sie auch noch mit Katie an!“
2	20:31	Leichte Untersicht	Halbnah	Ellen nimmt gerade ihre Ohrringe ab und entgegnet:	„Weil Patty sich Sorgen um sie macht!“
3	20:33	Normalsicht	Großaufnahme	David dreht sich weg:	„Schwachsinn!“
4	20:34	Leichte Untersicht	Halbnah	Ellen wendet wütend ein:	„Du hast dich schon aufgeführt, als wolltest du nicht dort sein, lange bevor Patty…
5	20:38	Normalsicht	Nahe	David hängt gerade sein Sakko in den Schrank	… Katie erwähnt hat!“
				David bekräftigt:	„So war’s ja auch.“
6	20:39	Leichte Untersicht	Halbnah	Ellen sagt:	„Ja, das hast du auch verdammt deutlich…
7	20:41	Normalsicht	Nahe	Blick auf David	… gezeigt!“
				Die Kamera fährt mit David nach links, während er spricht:	„Hör zu, es tut mir leid wenn ich mich nicht für…
				David geht an Ellen vorbei nach rechts, sie sind	… internationale

				jetzt beide im Bild	Aktienarbitragen…
8	20:46	Leichte Draufsicht	Halbtotale	Blick vom Badezimmer aus durch die geöffnete Tür zur Kommode, vor der Ellen und David stehen und streiten	… interessiere!“
				Ellen entgegnet: (Phil ist Pattys Ehemann, David arbeitet in einem Krankenhaus)	„Phil wollte sich nur unterhalten! Und als sich Patty nach dem Krankenhaus erkundigte, …“
				David unterbricht sie:	„… habe ich ihre Fragen beantwortet!“
9	20:51	Normalsicht	Nahe	Schnitt auf Ellen und David. Ellen widerspricht energisch:	**„Nein, du hast gegrunzt wie ein Höhlenmensch!“**
				David entgegnet widerspenstig:	„Ich hab’ ihr geantwortet, oder?“
				Ellen fühlt sich blamiert:	„Ich musste mich für dich entschuldigen!“
				David sieht sie ungläubig an:	„Du hast was? Oh!“
10	21:01	Leichte Draufsicht	Halbtotale	Wieder der Blick vom Badezimmer aus, in welches David nun geht. Ellen folgt ihm und sagt:	„So kannst du dich nicht benehmen, David. Patty ist das aufgefallen! Solche Essen gehören zu meinem Job – und zwar immer wieder!“

			Nahe	Die Kamera fährt näher an Ellen und David, dieser bleibt stur:	„Schon klar! Patty ist dein Boss – was aber nicht heißt, dass ich sie mögen muss!“
				Im Hintergrund klingelt ein Handy	
11	21:18	Leichte Draufsicht	Halbtotale	Ellen geht auf das Handy am Nachttisch neben dem Bett zu, setzt sich und hebt ab:	„Hallo?“
12	21:25	Leichte Draufsicht	Halbnah	David putzt sich die Zähne, wie im Badezimmerspiegel zu sehen ist. Er hört auf und sieht nach links durch die geöffnete Tür zu Ellen.	
13	21:26	Leichte Draufsicht	Halbtotale	Ellen sitzt regungslos und lautlos auf dem Bett.	

Damages – „Warum sollten wir das nicht?“ (Überraschung, Neugierde)

Beginn: 32.42 – Ende: 34:56

Nr.	Zeit	Kameraperspektive	Einstellungsgröße	Inhalt	Dialog
1	32:42	Leichte Draufsicht	Detailaufnahme	Blick auf eine Zeitung, den „New York Courier“. Auf der Titelseite prangt in großen Lettern: „Arthur Frobisher accused of Assault“, daneben ein großes Foto von Frobisher.	

				Als Textinsert in der deutschsprachigen Fassung ist zu lesen: „Milliardär flippt aus – Arthur Frobisher der Körperverletzung beschuldigt“	
		Normalsicht	Medium Shot	Die Kamera schwenkt langsam nach links oben, währenddessen kommt Patty Hewes (Ellens Chefin) von hinten ins Bild, man hört Ellens Stimme, die vorliest:	„Milliardär flippt aus – Arthur Frobisher der Körperverletzung beschuldigt.“
2	32:50	Normalsicht	Großaufnahme	Ellen sieht Patty an und fragt:	„Werden sie ihn festnehmen?“
3	32:53	Normalsicht	Medium Shot	Patty steht nach vorne gebeugt, sieht die Zeitung an und sagt zu Ellen:	„Einfache Körperverletzung ist nur ein Vergehen. Er kommt mit einer Geldstrafe davon. Aber eine solche Publicity kommt uns natürlich nur zugute. Gehen Sie schon?“
				Ellen erwidert:	„Nein, David arbeitet, also bleibe ich heute länger.“
				Patty sieht sie skeptisch an und fragt:	„Ich hab’ über David und Sie nachgedacht. Wollen Sie einen Drink?“
4	33:09	Normalsicht	Großaufnahme	Ellen sieht Patty an und lächelt unsicher:	„Gern.“

5	33:11	Normalsicht	Medium Shot	Patty lächelt Ellen an und dreht sich um in Richtung Tür	
6	33:12	Normalsicht	Großaufnahme	Ellen steht auf und geht hinter Patty her	
7	33:14	Leichte Draufsicht	Halbtotale	Patty geht aus dem Büro nach vorne durch einen Gang an der Kamera vorbei und sagt zu Ellen, die ihr folgt:	„Ich hoffe, Sie beide schaffen das.“
			Großaufnahme	Die Kamera bleibt auf Ellens Gesicht stehen als sie überrascht fragt:	**„Warum sollten wir das nicht?“**
			Amerikanische Einstellung	Die Kamera schwenkt weiter auf Ellen und Patty, die eine Schranktür öffnet und sagt:	„Es ist verdammt hart…
8	33:21	Leichte Untersicht	Nahe	Ellen steht neben einer niedrigen Lampe, Patty spricht weiter:	… Ellen, Mann und Karriere unter einen Hut zu bringen.“
9	33:24	Leichte Draufsicht	Halbnah	Patty hält Ellen grinsend einen Drink entgegen:	„Cheers!“
10	33:26	Leichte Untersicht	Nahe	Ellen nimmt Patty lächelnd das Glas aus der Hand, sieht hinein und sagt ebenfalls:	„Cheers!“
11	33:28	Leichte Draufsicht	Halbnah	Patty hat nun auch ein Glas in der Hand, Ellen nimmt einen Schluck während Patty erzählt:	„Wissen Sie – mein Freund auf dem College war der liebenswerteste, treueste Junge, den man sich vorstellen kann. Wir blieben nach dem Abschluss zusammen und er hat mir durch

					das Jurastudium geholfen.“
12	33:43	Leichte Untersicht	Nahe	Patty sieht gedankenverloren ins Leere, Ellen sitzt neben ihr und sieht sie an. Patty spricht weiter:	„Er hatte zwei Jobs, ganz undankbare, niedere…
13	33:47	Leichte Draufsicht	Halbnah	Schnitt auf Patty	… Tätigkeiten. Und… er war meine erste große Liebe.“
				Patty lächelt Ellen an, dann erzählt sie:	„Ich habe ihn nach dem Examen abserviert.“
14	33:54	Leichte Untersicht	Nahe	Ellen sitzt neben Patty, die einen Schluck nimmt, und fragt sie:	„Warum?“
15	33:56	Leichte Draufsicht	Halbnah	Schnitt auf Patty, die eiskalt erwidert:	„Er hatte keinerlei Ehrgeiz. Das hat mich angewidert.“
16	34:00	Leichte Untersicht	Nahe	Patty sitzt nickend neben Ellen, die etwas fassungslos fragt:	„Im Ernst?“
17	34:02	Leichte Draufsicht	Halbnah	Patty sieht Ellen mit einem Lächeln an	
18	34:04	Leichte Untersicht	Nahe	Ellen sieht Patty ungläubig an	
19	34:06	Leichte Draufsicht	Halbnah	Patty antwortet:	„Das ist Biologie, Ellen! Ein Mann muss das Alphatier sein wollen. Und der Trick dabei ist, ihm das Gefühl zu geben, dass er es ist.“

20	34:15	Leichte Untersicht	Nahe	Ellen nickt und fragt: (Phil ist Pattys Ehemann)	„Hat Phil deswegen beim Essen so viel geredet?“
21	34:18	Leichte Draufsicht	Halbnah	Patty schweigt erst, dann lacht sie	
22	34:22	Leichte Untersicht	Nahe	Ellen sieht sie etwas verständnislos an, mit einem halben Lächeln, Patty lacht weiter	
23	34:24	Leichte Draufsicht	Halbnah	Patty meint lachend:	„Ja, er hält gerne …
24	34:26	Leichte Untersicht	Nahe	Ellen sieht nachdenklich nach unten	… Hof!“
				Patty blickt Ellen an und fragt:	„Was?“
				Ellen erklärt:	„Ich habe David nicht erzählt, dass ich mich mit Gregory…
25	34:35	Leichte Draufsicht	Halbnah		… getroffen hab’.“
				Patty sieht sie an:	„Wirklich?“
26	34:38	Leichte Untersicht	Nahe	Ellen erklärt unsicher:	„Ich schätze, ich wollte ihn nicht beunruhigen.“
27	34:41	Leichte Draufsicht	Halbnahe	Patty erwidert:	„Ellen… Die meisten Männer können nicht mit ehrgeizigen Frauen umgehen.“
28	34:47	Leichte Untersicht	Großaufnahme	Die Kamera fährt näher an Ellens Gesicht, von Patty ist nun nur mehr seitlich der Kopf zu sehen. Patty sagt:	„Es wird vielleicht eine Weile…
29	34:50	Leichte Draufsicht	Großaufnahme	Schnitt auf Pattys Gesicht, Ellen ist nur mehr	… dauern, aber bemühen Sie sich,

				zum Teil im Bild	einen zu finden, der es kann.“
30	34:54	Leichte Untersicht	Großaufnahme	Ellen sieht Patty an, beide nicken kaum merklich.	

Damages – „David!“ (Angst)

Beginn: 01:13 – Ende: 01:28

Nr.	Zeit	Kameraperspektive	Einstellungsgröße	Inhalt	Dialog
1	01:13	Leichte Draufsicht	Halbtotale	Ellen Parsons schließt hektisch die Tür hinter sich, dreht sich um und schaut in die Wohnung. Die Kamera fährt auf sie zu	
			Amerikanische Einstellung	Ellen schreit in den Raum:	**„David!“**
2	01:14	Draufsicht	Detailaufnahme	Blick auf den Fußboden, Schwenk hin zu einer blutigen kleinen Imitation der Freiheitsstatue	
3	01:15	Leichte Draufsicht	Detailaufnahme	Schwenk nach oben	
			Medium Shot	Ellen kniet und sieht die Statuette an	
4	01:17	Leichte Draufsicht	Establishing Shot	Schnelle Kamerafahrt durch die offene Tür ins Badezimmer	
		Leichte Untersicht		Ellen steht vor der Badewanne, in der rechten	

				Hand hält sie die blutige Statuette, vor ihr in der Wanne liegt ihr Verlobter David Connor, ein Bein hängt heraus. Davids Augen sind offen, sein Kopf ist blutverschmiert, auch an Badewanne und Wand klebt Blut	
5	01:18	Untersicht	Detailaufnahme	Blick auf Davids blutige Hand, die ebenfalls aus der Badewanne hängt. Im Hintergrund ist verschwommen sein blutiges Gesicht zu erkennen	
6	01:19	Leichte Untersicht	Nahe	Schnitt auf Davids Gesicht und sein weißes, blutbeflecktes T-Shirt	
7	01:19	Leichte Untersicht	Detailaufnahme	Ellen lässt die blutige Statuette aus ihrer Hand fallen	
8	01:20	Normalsicht	Nahe	Blick von hinten auf Ellen, die vor der Wanne steht – gerade fällt die kleine Statue zu Boden	
9	01:21	Draufsicht	Halbtotale	Blick von der Zimmerdecke auf die Badewanne, Ellen stürzt zu ihrem toten Verlobten nieder	
10	01:23	Normalsicht	Nahe	Ellen nimmt Davids Kopf in ihre Arme und hält ihn fest	

CSI: Miami – „Ich ruf' da jetzt nicht an!" (Überraschung, Abscheu/Ekel)

Beginn: 06:18 – Ende: 07:07

Nr.	Zeit	Kameraperspektive	Einstellungsgröße	Inhalt	Dialog
1	06:18	Draufsicht	Detailaufnahme	Kreisrunder Blick auf gelbes Material	
2	06:21	Normalsicht	Nah	Tim Speedle sieht in Mikroskop	
3	06:22	Leichte Draufsicht	Establishing Shot	Eric Delko betritt das Labor, geht auf Speedle (der rechts im Bild sitzt) zu und spricht ihn an	„Es ist ein Lamborghini."
4	06:24	Leichte Draufsicht	Halbtotale	Kamera fährt nach unten und auf Delko und Speedle zu	
5	06:27	Leichte Untersicht	Medium Shot	Delko spricht weiter	„Der Farbcode ist 0056."
6	06:30	Normalsicht	Nahe	Speedle sieht sein Gegenüber an, von dem nur ein Teil des Arms im Bild ist, und sagt:	„Hellgelb metallic"
6	06:32	Leichte Untersicht	Nahe	Blick auf Delko, der sich überrascht zeigt:	„Genau!"
7	06:33	Normalsicht	Nahe	Schnitt auf Speedle:	„Was denkst du, wie viele Leute hier so ein Auto fahren?"
8	06:36	Leichte Untersicht	Nahe	Wieder Blick auf Delko:	„Vielleicht hat er's gemietet. Der kostet ‚ne Viertelmillion Dollar, da muss ‚ne alte Frau lang für stricken."
9	06:40	Normalsicht	Nahe	Speedle sieht auf etwas rechts von ihm (nicht	„Naja, das schränkt die Suche

				im Bild zu sehen):	ein.“
10	06:42			Speedle hebt den Telefonhörer auf:	„Hast du noch Kontakt zu der Maus, die bei der Zulassung arbeitet?“
11	06:46	Normalsicht	Großaufnahme	Delko sieht überrascht aus:	„Gina? Ja, wieso?“
12	06:47	Normalsicht	Nahe	Speedle reicht Delko den Telefonhörer:	„Hier!“
13	06:49	Leichte Untersicht	Großaufnahme	Delko verzieht das Gesicht:	„**Ich ruf’ da jetzt nicht an!** Das ist problematisch!“
14	06:51	Normalsicht	Großaufnahme	Speedle hört Delko zu, der sich beschwert:	„Wenn ich da jetzt anrufe muss ich mir wieder anhören, warum ich mich nicht gemeldet habe.“
15	06:55	Leichte Untersicht	Nahe	Delko sieht Speedle an, dessen Hand mit Telefonhörer unten im Bild zu sehen ist, und sagt:	„Darauf hab’ ich keine Lust!“
16	06:57	Normalsicht	Großaufnahme	Speedle gibt nicht nach:	„Eric, ab und zu musst du auch mal mit…
17	06:58	Leichte Untersicht	Nahe	Blick auf Delko, Speedle spricht weiter:	… denen ausgehen. Beziehungen muss man pflegen!“
18	07:00	Normalsicht	Großaufnahme	Schnitt auf Speedle:	„Du rufst sie jetzt an! Ich wähl’ auch für dich.“
19	07:04	Leichte Untersicht	Nahe	Delko verzieht wieder das Gesicht, presst die	„Toll.“

				Lippen aufeinander, seufzt, nimmt den Telefonhörer in die Hand und murmelt schlechtgelaunt:	

CSI: Miami – „Ich würde zu gern wissen, wie die aus deiner Wohnung hierher gekommen ist.“ (Überraschung, Freude)

Beginn: 25:15 – Ende: 26:13

Nr.	Zeit	Kameraperspektive	Einstellungsgröße	Inhalt	Dialog
1	25:15	Leichte Draufsicht	Supertotale	Blick auf einen Hafen, ein Schiff fährt auf die Kamera zu, sie schwenkt und fährt ihrerseits auf das Schiff zu. Als Hintergrundmusik ist rockige Gitarrenmusik zu hören.	
2	25:27	Normalsicht	Detailaufnahme	Es sind Beine zu sehen, die durch Müllberge waten, außerdem greifen Hände in Plastikhandschuhen nach dem Müll und lassen ihn wieder fallen	
3	25:29	Normalsicht	Amerikanische Einstellung	Eric Delko hebt im Schutzanzug Müll auf und wirft ihn wieder weg, die Kamera schwenkt nach links unten, wo Tim Speedle das Gleiche tut	

4	25:32	Leichte Draufsicht	Totale	Blick von oben auf die beiden CSI-Mitarbeiter auf dem Schiff, im Hintergrund sind der Hafen zu sehen sowie Wolkenkratzer und große Schiffe; ein Schiffshorn ist zu hören	
5	25:36	Normalsicht	Amerikanische Einstellung	Speedle sucht den Müll ab, Schwenk nach rechts auf Delko, der sich nach unten beugt und ausruft:	„Hey!“
6	25:39	Leichte Untersicht	Halbnah	Speedle sieht nach unten in Richtung Delko, der aus dem Hintergrund ruft:	„Sieh dir das hier mal an!“
7	25:41	Normalsicht	Detailaufnahme	Delko hebt etwas auf – eine tote Ratte kommt groß ins Bild	
8	25:43	Leichte Untersicht	Halbnah	Speedle blickt auf	
9	25:44	Leichte Untersicht	Halbnah	Delko sieht überrascht und erfreut aus über den Fund, er grinst hämisch und sagt:	**„Ich würde zu gern wissen, wie die aus deiner Wohnung hierher gekommen ist.“**
10	25:48	Leichte Untersicht	Halbnah	Speedle entgegnet mit genervtem Gesichtsausdruck:	„Haha, witzig.“
11	25:50	Leichte Untersicht	Halbnah	Delko steht mit ausgestreckter Hand da, wirft die Ratte aber doch weg	
12	25:51	Leichte Untersicht	Halbnah	Speedle sucht weiter, im Vordergrund fliegt die Ratte durchs Bild. Speedle seufzt:	„Puh.“

13	25:54	Normalsicht	Totale	Blick von hinten auf den Müllberg mit den suchenden Speedle und Delko, rechts begrenzt das Schiffsdeck die Sicht. Im Vordergrund öffnet sich gerade eine Zugbrücke, dahinter ist die Silhouette der Großstadt zu erkennen. Man hört das Läuten der Sicherheitsschranken der Brücke sowie weiterhin die Gitarrenmusik.	
14	25:59	Leichte Untersicht	Amerikanische Einstellung	Die Kamera fährt mit als sich Speedle nach unten beugt	
15	26:00	Normalsicht	Detailaufnahme	Speedle ist links im Bild zu sehen, seine Hände wühlen sich durch den Müll	
16	26:02	Normalsicht	Halbnah	Kamera schwenkt mit als Speedle etwas aus dem Müll zieht	
17	26:05	Normalsicht	Nahe	Speedle zieht weiter, er hat plötzlich ein schmerzverzerrtes Gesicht	
18	26:06	Leichte Untersicht	Amerikanische Einstellung	Speedle greift sich mit der rechten Hand auf den Rücken, gleichzeitig hält er etwas in der linken Hand, er jammert:	„Au!“
19	26:07	Leichte Untersicht	Halbnah	Im Vordergrund ist verschwommen Speedle mit dem Gegenstand in der Hand zu sehen, im Hintergrund rechts steht Delko. Die Kamera	

				schwenkt mit Delkos Bewegungen, der nach vorne links zu Speedle geht. Speedle gibt Schmerzenslaute von sich.	
				Delko streckt die Hände nach dem Gegenstand aus und fragt:	„Ist es das?“
20	26:10	Leichte Untersicht	Detailaufnahme	Blick auf die Hände, während der Gegenstand übergeben wird, der groß im Bild bleibt. Speedle antwortet:	„Ja, das ist unser Airbag.“

CSI: Miami – „Und wahrscheinlich ist es das, was mir ein Überschreiten der Grenze unmöglich macht.“ (Traurigkeit, Vertrauen)

Beginn: 18:22 – Ende: 20:15

Nr.	Zeit	Kameraperspektive	Einstellungsgröße	Inhalt	Dialog
1	18:22	Leichte Untersicht	Establishing Shot	Kamera schwenkt nach links Horatio Caine und Yelina Salas gehen aufeinander zu, sie bleiben im hinteren Bildteil stehen	
2	18:26			Caine lacht, Salas sieht ihn an und sagt:	„Ich vermute, das State Department hält die Hand über Clavo.“

				Man hört Klaviermusik im Hintergrund, Caine antwortet:	„Nationale Sicherheit…“
3	18:31	Leichte Untersicht	Amerikanische Einstellung	Salas fragt:	„Was jetzt? Hören wir auf?“
				Caine:	„Nein, Was ist schon von großer Dauer? Aus Verbündeten werden ganz schnell Feinde. Und dann sind Beweise wieder wichtig.“
				Salas:	„Also weiter.“
				Caine:	„Mhm… Ist das nicht Statler da drüben?“
4	18:44	Leichte Untersicht	Medium Shot	Blick auf einen Mann im Nebenraum, der an einem Bartresen sitzt; Salas:	„Ja, er meinte…
5	18:47	Normalsicht	Over-the-Shoulder Close-Up	Salas blickt Caine an und spricht weiter:	… du hättest gesagt, er solle mit…
6	18:48	Normalsicht	Over-the-Shoulder Close-Up	Caine sieht Salas an, sie redet weiter:	… mir reden.“
				Caine antwortet:	„Nein, das ist nicht wahr. Ich habe ihm gesagt, ich kann nicht

					für dich sprechen.“
7	18:52	Normalsicht	Over-the-Shoulder Close-Up	Salas blickt Caine an:	“Für dich aber auch nicht.”
8	18:53	Normalsicht	Over-the-Shoulder Close-Up	Die Klaviermusik geht langsam über in traurige, langsame Stimmungsmusik. Caine holt tief Luft und sagt:	„Na gut. Ich denke ständig, wenn ich dich vor Raymond kennengelernt hätte, wären wir vielleicht…“
				Salas unterbricht ihn:	„Raymond ist tot.“
				Caine blickt Salas traurig an und spricht weiter:	„Ja.“
	19:07			Caine:	**„Und wahrscheinlich ist es das, was mir ein Überschreiten der Grenze unmöglich macht.“**
9	19:11	Normalsicht	Großaufnahme	Salas sieht leicht nach oben und blickt Caine an:	„Richtig. Ich bin die Frau deines toten Bruders, hm?“
10	19:17	Normalsicht	Großaufnahme	Caine sieht sie nur wortlos an	
11	19:19	Normalsicht	Großaufnahme	Salas erwidert den Blick und sagt:	„Du kannst nichts für Raymonds Tod. Wir beide nicht. Du hast ihn mehr als jeder andere ständig gewarnt, vorsichtig …
12	19:25	Normalsicht	Großaufnahme	Schnitt auf Caine	… zu sein.“
				Caine entgegnet:	„Ich hätte nachdrücklicher sein

					müssen.“
13	19:28	Normalsicht	Großaufnahme	Salas kontert:	„Das ist nicht wahr! Das stimmt doch nicht.“
14	19:32	Normalsicht	Großaufnahme	Caine sieht sie an und seufzt lediglich:	„Hm.“
15	19:34	Normalsicht	Großaufnahme	Salas sagt sanfter:	„Aber ich weiß, was du…
16	19:37	Normalsicht	Großaufnahme	Blick auf Caine	… meinst.“
17	19:40	Normalsicht	Großaufnahme	Salas sieht Caine an und sagt:	„Okay…“
18	19:43	Normalsicht	Großaufnahme	Caine blickt sie wortlos an	
19	19:46	Normalsicht	Großaufnahme	Salas kapituliert:	„Ich nehme an, ich muss die Dinge im Moment so nehmen, wie sie sind.“
20	19:51	Normalsicht	Großaufnahme	Caine nickt kaum merklich und lächelt traurig	
21	19:52	Normalsicht	Over-the-Shoulder Close-Up	Salas spricht weiter, mit sarkastischem Unterton:	„Aber ich gehe nicht mit Statler aus. Dieses Grinsen…“
22	19:58	Leichte Untersicht	Medium Shot	Wieder Schnitt auf den Mann am Tresen, Caine bemerkt dazu nur:	„Mhm.“
23	20:00	Leichte Untersicht	Amerikanische Einstellung	Salas und Caine stehen einander gegenüber, Caine merkt trocken an:	„Das ist mir auch aufgefallen.“
				Salas lächelt:	„Aham…“
24	20:03	Leichte Untersicht	Over-the-Shoulder Close-Up	Caine sieht Salas an und sagt:	„Übrigens, jetzt…

25	20:05	Normalsicht	Nahe	Salas blickt Caine an	… tut er’s grade wieder.“
				Salas verdreht die Augen und lächelt	
26	20:06	Leichte Untersicht	Over-the-Shoulder Close-Up	Caine geht nach hinten weg und sagt:	„Wir sehen uns.“
27	20:07	Leichte Untersicht	Amerikanische Einstellung	Salas erwidert:	„Ja, klar.“
				Salas bleibt stehen, Caine sagt im Hintergrund noch zu Restaurantangestellten:	„Dankeschön, wiedersehen.“
				Salas fährt sich durch die Haare, die Rezeptionistin verabschiedet sich von Caine:	„Schönen Abend noch.“
				Salas dreht sich nach links um und geht nach rechts aus dem Bild	

universität wien

Herzlich willkommen!

Vielen Dank für Ihre Mithilfe bei meiner Diplomarbeit zum Thema

"Stimme im Fernsehen - Auditive Emotionswahrnehmung im audiovisuellen Medium"

Bitte beachten Sie die folgenden Instruktionen:

Aus technischen Gründen funktioniert der Fragebogen nur mit dem Mozilla Firefox - **nicht dem Internet Explorer**.

Die Beantwortung der Fragen nimmt etwa **10 bis 15 Minuten** in Anspruch. Bitte beginnen Sie mit der Befragung nur, wenn Sie diese Zeit **am Stück, alleine und in ruhiger Umgebung** aufbringen können. Ansonsten speichern Sie bitte den Link in Ihrem Webbrowser und nehmen zu einem späteren Zeitpunkt an der Befragung teil. Wenn Sie die Fragen nicht wahrheitsgemäß beantworten, in Eile wahllos Antworten anklicken oder den Fragebogen abbrechen, ist dieser für meine Diplomarbeit leider nicht verwendbar. Das Gleiche gilt, wenn Sie den Fragebogen nicht allein bearbeiten.

Bitte benützen Sie **niemals den "Zurück"-Button** Ihres Webbrowsers, da eine Auswertung aus technischen Gründen sonst nicht möglich ist.

Bitte antworten Sie auf alle Fragen spontan, aus dem Bauch heraus. Es gibt kein "falsch" und kein "richtig"! Für mich und meine Arbeit zählt lediglich Ihre Meinung :-)

Meine Umfrage beinhaltet Fragebögen und einen Multimedia-Teil. Für letzteren benötigen Sie unbedingt **Sound** - überprüfen Sie bitte vor dem Start der Befragung, dass dieser aktiviert und laut genug eingestellt ist.

Ihre Daten werden natürlich vertraulich behandelt und nicht (!) weitergegeben. Der Fragebogen ist **anonymisiert**, das heißt, niemand kann feststellen, wer welche Antworten gegeben hat. Bitte beantworten Sie daher auch die Fragen zu Ihren persönlichen Daten - der Fragebogen ist für mich sonst leider nicht auswertbar.

Bitte senden Sie den Link an so viele Menschen wie möglich, er ist auf Personen von 16 Jahren aufwärts abgestimmt - ich freue mich über rege Beteiligung aller Altersklassen. Vielen Dank für Ihre Mithilfe!

Dankeschön und viel Vergnügen,
Bernadette Geißler

Befragung starten

Stimmungserhebung (Teil 1 von 6)

Wie stark empfinden Sie diese Gefühle im Moment?	Gar nicht	Wenig	Mittelmäßig	Ziemlich	Sehr stark
Freude	○	○	○	○	○
Vertrauen	○	○	○	○	○
Angst	○	○	○	○	○
Überraschung	○	○	○	○	○
Traurigkeit	○	○	○	○	○
Ekel	○	○	○	○	○
Wut	○	○	○	○	○
Neugierde	○	○	○	○	○

Weiter

Fragebogen (Teil 2 von 6)

Phantasie

Frage	Trifft gar nicht zu	Trifft nicht zu	Trifft teilweise zu	Trifft zu	Trifft sehr zu
Ich finde es etwas übertrieben, sich in Kinofilme oder Fernsehserien hineinzusteigern	○	○	○	○	○
In die Gefühle von Romanfiguren lebe ich mich richtig hinein	○	○	○	○	○
Ich neige dazu, Kinofilme derart mitzuerleben, dass ich empfinde, als wäre ich selbst eine der handelnden Personen	○	○	○	○	○
Ich neige dazu, Fernsehserien derart mitzuerleben, dass ich empfinde, als wäre ich selbst eine der handelnden Personen	○	○	○	○	○
Bei einem interessanten Roman stelle ich mir vor, wie ich in so einer Situation zurecht käme	○	○	○	○	○
Es passiert mir eher selten, in einem guten Kinofilm besonders aufzugehen	○	○	○	○	○
Es passiert mir eher selten, in einer guten Fernsehserie besonders aufzugehen	○	○	○	○	○
Bei einem guten Kinofilm stelle ich mir vor, wie ich wohl in so einer Situation zurecht käme	○	○	○	○	○
Bei einer guten Fernsehserie stelle ich mir vor, wie ich wohl in so einer Situation zurecht käme	○	○	○	○	○

Reale Situationen

Frage	Trifft gar nicht zu	Trifft nicht zu	Trifft teilweise zu	Trifft zu	Trifft sehr zu

Reale Situationen

Frage	Trifft gar nicht zu	Trifft nicht zu	Trifft teilweise zu	Trifft zu	Trifft sehr zu
Der Anblick weinender Menschen bringt mich aus der Fassung	○	○	○	○	○
Ich neige dazu, mich in die Probleme eines Freundes hineinzuleben	○	○	○	○	○
Wenn ich einen sehr alten Menschen sehe, frage ich mich oft, wie ich mich an seiner/ihrer Stelle fühlen würde	○	○	○	○	○
Filme über Krieg und Töten regen mich innerlich auf	○	○	○	○	○
Es beunruhigt mich mehr als die meisten Menschen, wenn ich sehe, wie sich ein Freund verletzt	○	○	○	○	○
Ich bin oft ziemlich berührt durch Dinge, die vor meinen Augen passieren	○	○	○	○	○
Ich fühle oft Betroffenheit und Mitgefühl mit anderen Menschen, die weniger glücklich sind als ich	○	○	○	○	○
Missgeschicke anderer Menschen berühren mich meist nicht sehr	○	○	○	○	○
Manchmal versuche ich meine Freunde besser zu verstehen, indem ich mir die Dinge aus ihrer Sicht vorstelle	○	○	○	○	○
Wenn ich ein behindertes Kind sehe, versuche ich mir vorzustellen, wie es sich in bestimmten Situationen fühlt	○	○	○	○	○
Es macht mich traurig, in einer Gruppe einen einsamen Menschen zu sehen	○	○	○	○	○
Die Menschen um mich haben einen großen Einfluss auf meine Stimmung	○	○	○	○	○

Weiter

Serienkenntnis-Erhebung (Teil 3 von 6)

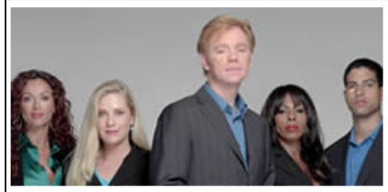
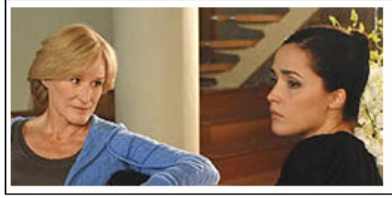
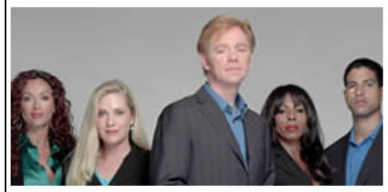
Bild	Schätzen Sie: Wie oft haben Sie diese Fernsehserie (auf Deutsch) INSGESAMT bereits gesehen?	Noch nie	1 Mal	2 bis 5 Mal	5 bis 10 Mal	Öfter als 10 Mal
	CSI: Miami	○	○	○	○	○
	Damages - Im Netz der Macht	○	○	○	○	○

Bild	Sehen Sie sich diese Serie REGELMÄSSIG (auf Deutsch) im Fernsehen an? Bzw. haben Sie die Serie REGELMÄSSIG (auf Deutsch) verfolgt, als sie im TV zu sehen war?	Verfolg(t)e ich nicht regelmäßig	1 Mal pro Monat	2 bis 3 Mal pro Monat	1 Mal pro Woche	Öfter als 1 Mal pro Woche
	CSI: Miami	○	○	○	○	○
	Damages - Im Netz der Macht	○	○	○	○	○

Weiter

Video-Fragebogen (Teil 4 von 6)

Bitte beachten Sie die Instruktionen:

- Überprüfen Sie, dass der Sound Ihres Computers aktiviert ist bzw. die Lautsprecherboxen Ihres Computers eingeschaltet sind.
- Warten Sie, bis die Videos geladen sind (kann einige Sekunden dauern).
- Wundern Sie sich nicht: Die Videosequenzen sind sehr kurz. Es gibt Videos mit Bild und Ton, Videos mit schwarzem Bild (nur Ton) sowie Videos ohne Ton (nur Bild).
- Sehen Sie sich ein Video an (wenn Sie möchten auch mehrmals - ich empfehle jedoch nicht öfter als 3 Mal pro Video).
- Klicken Sie nun an, welches Gefühl Sie in der Szene hören/sehen. Klicken Sie bitte **mindestens ein, maximal zwei** Gefühl/e an. Einige Szenen vermitteln Ihnen vielleicht nur ein Gefühl, andere zwei verschiedene. Auch wenn Sie unsicher sind: klicken Sie bitte mindestens ein Gefühl an - jenes, das Ihrer Meinung nach am besten zutrifft. Es gibt kein "falsch" und kein "richtig"!
- Klicken Sie bitte **nach jeder Gefühlseinschätzung auf "Speichern"** und gehen zum nächsten Videobeispiel über.

Beispiel 1 (nur Bild, kein Ton)

Was denken Sie: Welches Gefühl vermittelt der/die SprecherIn/SchauspielerIn?

- ☐ Freude
- ☐ Vertrauen
- ☐ Angst
- ☐ Erstaunen
- ☐ Traurigkeit
- ☐ Ekel/Abscheu

☐ Wut
☐ Neugierde
Speichern

Beispiel 2 (Bild & Ton)

Was denken Sie: Welches Gefühl vermittelt der/die SprecherIn/SchauspielerIn?

☐ Freude
☐ Vertrauen
☐ Angst
☐ Erstaunen
☐ Traurigkeit
☐ Ekel/Abscheu
☐ Wut
☐ Neugierde
Speichern

Beispiel 3 (nur Bild, kein Ton)

Beispiel 3 (nur Bild, kein Ton)

Was denken Sie: Welches Gefühl vermittelt der/die SprecherIn/SchauspielerIn?

- [ ] Freude
- [ ] Vertrauen
- [ ] Angst
- [ ] Erstaunen
- [ ] Traurigkeit
- [ ] Ekel/Abscheu
- [ ] Wut
- [ ] Neugierde

Speichern

Beispiel 4 (nur Ton, kein Bild)

Beispiel 4 (nur Ton, kein Bild)

Was denken Sie: Welches Gefühl vermittelt der/die SprecherIn/SchauspielerIn?

- ☐ Freude
- ☐ Vertrauen
- ☐ Angst
- ☐ Erstaunen
- ☐ Traurigkeit
- ☐ Ekel/Abscheu
- ☐ Wut
- ☐ Neugierde

Speichern

Beispiel 5 (Bild & Ton)

Beispiel 5 (Bild & Ton)

Was denken Sie: Welches Gefühl vermittelt der/die SprecherIn/SchauspielerIn?

- ☐ Freude
- ☐ Vertrauen
- ☐ Angst
- ☐ Erstaunen
- ☐ Traurigkeit
- ☐ Ekel/Abscheu
- ☐ Wut
- ☐ Neugierde

Speichern

Beispiel 6 (nur Bild, kein Ton)

Beispiel 6 (nur Bild, kein Ton)

Was denken Sie: Welches Gefühl vermittelt der/die SprecherIn/SchauspielerIn?

- ☐ Freude
- ☐ Vertrauen
- ☐ Angst
- ☐ Erstaunen
- ☐ Traurigkeit
- ☐ Ekel/Abscheu
- ☐ Wut
- ☐ Neugierde

Speichern

Beispiel 7 (nur Ton, kein Bild)

☐ Erstaunen
☐ Traurigkeit
☐ Ekel/Abscheu
☐ Wut
☐ Neugierde
Speichern

Beispiel 7 (nur Ton, kein Bild)

Was denken Sie: Welches Gefühl vermittelt der/die SprecherIn/SchauspielerIn?

☐ Freude
☐ Vertrauen
☐ Angst
☐ Erstaunen
☐ Traurigkeit
☐ Ekel/Abscheu
☐ Wut
☐ Neugierde
Speichern

Weiter

Video-Fragebogen (Teil 4 von 6)

Bitte beachten Sie die Instruktionen:

- Überprüfen Sie, dass der Sound Ihres Computers aktiviert ist bzw. die Lautsprecherboxen Ihres Computers eingeschaltet sind.
- Warten Sie, bis die Videos geladen sind (kann einige Sekunden dauern).
- Wundern Sie sich nicht: Die Videosequenzen sind sehr kurz. Es gibt Videos mit Bild und Ton, Videos mit schwarzem Bild (nur Ton) sowie Videos ohne Ton (nur Bild).
- Sehen Sie sich ein Video an (wenn Sie möchten auch mehrmals - ich empfehle jedoch nicht öfter als 3 Mal pro Video).
- Klicken Sie nun an, welches Gefühl Sie in der Szene hören/sehen. Klicken Sie bitte **mindestens ein, maximal zwei** Gefühl/e an. Einige Szenen vermitteln Ihnen vielleicht nur ein Gefühl, andere zwei verschiedene. Auch wenn Sie unsicher sind: klicken Sie bitte mindestens ein Gefühl an - jenes, das Ihrer Meinung nach am besten zutrifft. Es gibt kein "falsch" und kein "richtig"!
- Klicken Sie bitte **nach jeder Gefühlseinschätzung auf "Speichern"** und gehen zum nächsten Videobeispiel über.

Beispiel 1 (Bild & Ton)

Was denken Sie: Welches Gefühl vermittelt der/die SprecherIn/SchauspielerIn?

- ☐ Freude
- ☐ Vertrauen
- ☐ Angst
- ☐ Erstaunen
- ☐ Traurigkeit
- ☐ Ekel/Abscheu

☐ Wut
☐ Neugierde
Speichern

Beispiel 2 (nur Bild, kein Ton)

Was denken Sie: Welches Gefühl vermittelt der/die SprecherIn/SchauspielerIn?

☐ Freude
☐ Vertrauen
☐ Angst
☐ Erstaunen
☐ Traurigkeit
☐ Ekel/Abscheu
☐ Wut
☐ Neugierde
Speichern

Beispiel 3 (Bild & Ton)

Beispiel 3 (Bild & Ton)

Was denken Sie: Welches Gefühl vermittelt der/die SprecherIn/SchauspielerIn?

- ☐ Freude
- ☐ Vertrauen
- ☐ Angst
- ☐ Erstaunen
- ☐ Traurigkeit
- ☐ Ekel/Abscheu
- ☐ Wut
- ☐ Neugierde

Speichern

Beispiel 4 (nur Ton, kein Bild)

Beispiel 4 (nur Ton, kein Bild)

Was denken Sie: Welches Gefühl vermittelt der/die SprecherIn/SchauspielerIn?

- ☐ Freude
- ☐ Vertrauen
- ☐ Angst
- ☐ Erstaunen
- ☐ Traurigkeit
- ☐ Ekel/Abscheu
- ☐ Wut
- ☐ Neugierde

Speichern

Beispiel 5 (nur Bild, kein Ton)

Beispiel 5 (nur Bild, kein Ton)

Was denken Sie: Welches Gefühl vermittelt der/die SprecherIn/SchauspielerIn?

- ☐ Freude
- ☐ Vertrauen
- ☐ Angst
- ☐ Erstaunen
- ☐ Traurigkeit
- ☐ Ekel/Abscheu
- ☐ Wut
- ☐ Neugierde

Speichern

Beispiel 6 (Bild & Ton)

Beispiel 6 (Bild & Ton)

Was denken Sie: Welches Gefühl vermittelt der/die SprecherIn/SchauspielerIn?

- ☐ Freude
- ☐ Vertrauen
- ☐ Angst
- ☐ Erstaunen
- ☐ Traurigkeit
- ☐ Ekel/Abscheu
- ☐ Wut
- ☐ Neugierde

Speichern

Beispiel 7 (nur Ton, kein Bild)

☐ Erstaunen
☐ Traurigkeit
☐ Ekel/Abscheu
☐ Wut
☐ Neugierde
Speichern

Beispiel 7 (nur Ton, kein Bild)

Was denken Sie: Welches Gefühl vermittelt der/die SprecherIn/SchauspielerIn?

☐ Freude
☐ Vertrauen
☐ Angst
☐ Erstaunen
☐ Traurigkeit
☐ Ekel/Abscheu
☐ Wut
☐ Neugierde
Speichern

Weiter

Video-Fragebogen (Teil 4 von 6)

Bitte beachten Sie die Instruktionen:

- Überprüfen Sie, dass der Sound Ihres Computers aktiviert ist bzw. die Lautsprecherboxen Ihres Computers eingeschaltet sind.
- Warten Sie, bis die Videos geladen sind (kann einige Sekunden dauern).
- Wundern Sie sich nicht: Die Videosequenzen sind sehr kurz. Es gibt Videos mit Bild und Ton, Videos mit schwarzem Bild (nur Ton) sowie Videos ohne Ton (nur Bild).
- Sehen Sie sich ein Video an (wenn Sie möchten auch mehrmals - ich empfehle jedoch nicht öfter als 3 Mal pro Video).
- Klicken Sie nun an, welches Gefühl Sie in der Szene hören/sehen. Klicken Sie bitte **mindestens ein, maximal zwei** Gefühl/e an. Einige Szenen vermitteln Ihnen vielleicht nur ein Gefühl, andere zwei verschiedene. Auch wenn Sie unsicher sind: klicken Sie bitte mindestens ein Gefühl an - jenes, das Ihrer Meinung nach am besten zutrifft. Es gibt kein "falsch" und kein "richtig"!
- Klicken Sie bitte **nach jeder Gefühlseinschätzung auf "Speichern"** und gehen zum nächsten Videobeispiel über.

Beispiel 1 (nur Ton, kein Bild)

Was denken Sie: Welches Gefühl vermittelt der/die SprecherIn/SchauspielerIn?

- ☐ Freude
- ☐ Vertrauen
- ☐ Angst
- ☐ Erstaunen
- ☐ Traurigkeit
- ☐ Ekel/Abscheu

☐ Wut
☐ Neugierde
Speichern

Beispiel 2 (Bild & Ton)

Was denken Sie: Welches Gefühl vermittelt der/die SprecherIn/SchauspielerIn?

☐ Freude
☐ Vertrauen
☐ Angst
☐ Erstaunen
☐ Traurigkeit
☐ Ekel/Abscheu
☐ Wut
☐ Neugierde
Speichern

Beispiel 3 (nur Bild, kein Ton)

Beispiel 3 (nur Bild, kein Ton)

Was denken Sie: Welches Gefühl vermittelt der/die SprecherIn/SchauspielerIn?

- ☐ Freude
- ☐ Vertrauen
- ☐ Angst
- ☐ Erstaunen
- ☐ Traurigkeit
- ☐ Ekel/Abscheu
- ☐ Wut
- ☐ Neugierde

Speichern

Beispiel 4 (nur Ton, kein Bild)

Beispiel 4 (nur Ton, kein Bild)

Was denken Sie: Welches Gefühl vermittelt der/die SprecherIn/SchauspielerIn?

- ☐ Freude
- ☐ Vertrauen
- ☐ Angst
- ☐ Erstaunen
- ☐ Traurigkeit
- ☐ Ekel/Abscheu
- ☐ Wut
- ☐ Neugierde

Speichern

Beispiel 5 (Bild & Ton)

Beispiel 5 (Bild & Ton)

Was denken Sie: Welches Gefühl vermittelt der/die SprecherIn/SchauspielerIn?

- [ ] Freude
- [ ] Vertrauen
- [ ] Angst
- [ ] Erstaunen
- [ ] Traurigkeit
- [ ] Ekel/Abscheu
- [ ] Wut
- [ ] Neugierde

Speichern

Beispiel 6 (nur Bild, kein Ton)

Beispiel 6 (nur Bild, kein Ton)

Was denken Sie: Welches Gefühl vermittelt der/die SprecherIn/SchauspielerIn?

- [ ] Freude
- [ ] Vertrauen
- [ ] Angst
- [ ] Erstaunen
- [ ] Traurigkeit
- [ ] Ekel/Abscheu
- [ ] Wut
- [ ] Neugierde

Speichern

Beispiel 7 (nur Ton, kein Bild)

☐ Erstaunen
☐ Traurigkeit
☐ Ekel/Abscheu
☐ Wut
☐ Neugierde
Speichern

Beispiel 7 (nur Ton, kein Bild)

Was denken Sie: Welches Gefühl vermittelt der/die SprecherIn/SchauspielerIn?

☐ Freude
☐ Vertrauen
☐ Angst
☐ Erstaunen
☐ Traurigkeit
☐ Ekel/Abscheu
☐ Wut
☐ Neugierde
Speichern

Weiter

Persönlichkeitsfragebogen (Teil 5 von 6)

Frage	Stimmt	Stimmt nicht
Ich gehe abends gerne aus	○	○
Ich habe/hatte einen Beruf, der mich voll befriedigt	○	○
Ich habe fast immer eine schlagfertige Antwort parat	○	○
Ich würde kaum zögern, auch alte und schwerbehinderte Menschen zu pflegen	○	○
Ich würde mich beim Kellner oder Geschäftsführer eines Restaurants beschweren, wenn ein schlechtes Essen serviert wird	○	○
Ich bin unternehmungslustiger als die meisten meiner Bekannten	○	○
Ich habe manchmal ein Gefühl der Teilnahmslosigkeit und inneren Leere	○	○
Sind wir in ausgelassener Runde, so überkommt mich oft eine große Lust zu groben Streichen	○	○
Ich bin der Ansicht, die Menschen in den Entwicklungsländern sollten sich zuerst einmal selbst helfen	○	○
Ich bin ungern mit Menschen zusammen, die ich noch nicht kenne	○	○
Ich fühle mich über meine Familie hinaus für andere Menschen verantwortlich	○	○
Ich bin oft nervös, weil zu viel auf mich einströmt	○	○
Wenn ich noch einmal geboren würde, dann würde ich nicht anders leben wollen	○	○
Ich übernehme bei gemeinsamen Unternehmungen gerne die Führung	○	○
Ich lebe mit mir selbst in Frieden und ohne innere Konflikte	○	○
Ich finde, jeder Mensch soll sehen, wie er zurecht kommt	○	○
Ich habe gern mit Aufgaben zu tun, die schnelles Handeln verlangen	○	○
Ich denke oft, dass ich meinen Konsum einschränken müsste, um dann an benachteiligte Menschen abzugeben	○	○
Meine Familie und meine Bekannten können mich im Grunde kaum richtig verstehen	○	○
Ich kann in eine ziemlich langweilige Gesellschaft schnell Leben bringen	○	○
Wenn jemand weint, möchte ich ihn am liebsten umarmen und trösten	○	○
Bei wichtigen Dingen bin ich bereit, mit anderen energisch zu konkurrieren	○	○

Bei wichtigen Dingen bin ich bereit, mit anderen energisch zu konkurrieren	○	○
Ich fühle mich oft wie ein Pulverfass kurz vor der Explosion	○	○
Ich würde mich selbst eher als gesprächig bezeichnen	○	○
Termindruck und Hektik lösen bei mir körperliche Beschwerden aus	○	○
In meinem bisherigen Leben habe ich kaum das verwirklichen können, was in mir steckt	○	○
Ich bin ziemlich lebhaft	○	○
Ich mache mir oft Sorgen um meine Gesundheit	○	○
Es fällt mir schwer, vor einer großen Gruppe von Menschen zu sprechen oder vorzutragen	○	○
Ich bin im Grunde eher ein ängstlicher Mensch	○	○
Ich gebe gelegentlich Geld und Spenden für Katastrophenhilfe, Caritas oder andere wohltätige Organisationen	○	○
Ich bin häufiger abgespannt, matt und erschöpft	○	○
Oft habe ich alles gründlich satt	○	○
Ich schließe nur langsam Freundschaften	○	○
Manchmal habe ich ohne eigentlichen Grund ein Gefühl unbestimmter Gefahr oder Angst	○	○
Ich werde ziemlich leicht verlegen	○	○
Wenn mich ein Fremder um eine kleine Geldspende bittet, ist mir das ziemlich lästig	○	○
Ich bin immer guter Laune	○	○
Ich spiele anderen Leuten gerne einen harmlosen Streich	○	○
Ich erröte leicht	○	○
Ich bin selten in bedrückter, unglücklicher Stimmung	○	○
Bei Geselligkeiten und öffentlichen Veranstaltungen bleibe ich lieber im Hintergrund	○	○
Ich träume oft von Dingen, die doch nicht verwirklicht werden können	○	○
Ich grüble viel über mein bisheriges Leben nach	○	○
Ich nehme mir viel Zeit, anderen Menschen geduldig zuzuhören, wenn sie von ihren Sorgen erzählen	○	○
Oft rege ich mich zu rasch über jemanden auf	○	○

Ich bin mit meinen gegenwärtigen Lebensbedingungen oft unzufrieden	○	○
Beim Reisen schaue ich lieber auf die Landschaft als mich mit den Mitreisenden zu unterhalten	○	○
Da der Staat schon für Sozialhilfe sorgt, brauche ich im Einzelnen nicht zu helfen	○	○
Meine Laune wechselt ziemlich oft	○	○
Es fällt mir schwer, den richtigen Gesprächsstoff zu finden, wenn ich jemanden kennenlerne	○	○
Alles in allem bin ich ausgesprochen zufrieden mit meinem bisherigen Leben	○	○
Ich habe oft das Gefühl, im Stress zu sein	○	○
Meine Partnerbeziehung/Ehe ist gut	○	○
Ich habe schon unbezahlt beim Roten Kreuz, in anderen sozialen Einrichtungen oder in meiner Gemeinde geholfen	○	○
Meistens blicke ich voller Zuversicht in die Zukunft	○	○

Weiter

Persönliche Daten (Teil 6 von 6)

Geschlecht	○ weiblich ○ männlich	
Alter	Unter 16	
Höchster Schulabschluss	Pflichtschule Lehre, Fachschule Matura Universität, Fachhochschule	
Sind Sie berufstätig?	Ja Ja, und Student/in Hausfrau/Hausmann ***oder sind Sie*** Schüler/in Student/in	
Familienstand	ledig Freund/in, Lebenspartner/in verheiratet getrennt lebend/geschieden verwitwet	
Haben Sie Kinder (egal ob leiblich, adoptiert oder durch Partnerschaft/Heirat in der Familie)?	○ ja ○ nein	○ 1 Tochter ○ 2 Töchter ○ 3 oder mehr Töchter ○ Keine Tochter ○ 1 Sohn ○ 2 Söhne ○ 3 oder mehr Söhne ○ Kein Sohn
Sind Sie mit Geschwistern aufgewachsen?	○ ja ○ nein	○ 1 Schwester ○ 2 Schwestern ○ 3 oder mehr Schwestern ○ Keine Schwester ○ 1 Bruder ○ 2 Brüder ○ 3 oder mehr Brüder ○ Kein Bruder

Vielen Dank für Ihre Teilnahme!

Fragen oder Anregungen bitte per Email an geissler_bernadette[redacted]
Sie können das Fenster jetzt schließen.

Lebenslauf

Name:	Bernadette Geißler
Geburtstag und Geburtsort:	07. 07. 1983 in Salzburg
Nationalität:	Österreich
Familie:	Vater Josef, Mutter Elisabeth, Schwester Patricia

Schule und Studium:

1989 – 1993	Volksschule Ottensheim/OÖ
1993 – 1997	Bundesgymnasium Körnerschule Linz/OÖ
1997 – 2002	HBLA Schwerpunkt Kulturtouristik Linz-Auhof/OÖ
18. 6. 2002	Matura mit gutem Erfolg
WiSe 2002 – SoSe 2009	Studium der Publizistik & Kommunikationswissenschaften, in der Fächerkombination mit Theaterwissenschaft und Spanisch an der Universität Wien

Berufserfahrung:

04/2004 – 05/2006	Redakteurin CHiLLi.cc (Innenpolitik, Kultur, Gesellschaft)
02/2005 – 05/2006	Chefin vom Dienst (CvD) CHiLLi.cc
09/2005 – 05/2006	Stellvertretende Chefredakteurin CHiLLi.cc
05/2006 – 11/2008	Redakteurin Krone Multimedia (krone.at) (Innen- und Außenpolitik, Chronik, Gesellschaft, Kino, Computer- und Videospiele, Multimedia-Content)
07/2008 – 11/2008	Leiterin Aktuelle Nachrichten Krone Multimedia (krone.at)
Seit 06/2009	Redakteurin GamingXP (Computer- und Videospiele)