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.


„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 .................................................................................................................................................. 1
2. Theorie .................................................................................................................................................... 2
  2.1. Die Anfänge der Stimmforschung .............................................................................................. 2
  2.2. Physikalische Grundlagen ........................................................................................................ 3
  2.3. Physiologische Grundlagen ....................................................................................................... 4
    2.3.1. Emotionen .................................................................................................................................. 4
    2.3.1.1. Gehirn ............................................................................................................................... 4
    2.3.1.2. Biochemische Botenstoffe ............................................................................................... 6
    2.3.2. Voraussetzungen für Stimmbildung und Stimmerkennung .................................................. 6
      2.3.2.1. Gehirn und Nervensystem ............................................................................................... 6
      2.3.2.2. Atmung und Sprechen ..................................................................................................... 8
      2.3.2.3. Die Stimmproduktion ...................................................................................................... 9
      2.3.2.4. Das Gehör .......................................................................................................................... 10
    2.3.3. Stimme, Sprache und Gehör im körperlichen Entwicklungsprozess ..................................... 11
      2.3.3.1. Stimme .............................................................................................................................. 11
      2.3.3.2. Sprache ............................................................................................................................ 12
      2.3.3.3. Gehör .............................................................................................................................. 13
    2.3.4. Die Besonderheiten der Stimme ............................................................................................ 14
      2.3.4.1. Stimmlage und -lautstärke ............................................................................................... 15
      2.3.4.2. Stimmgattungen ............................................................................................................... 15
      2.3.4.3. Stimmklang ...................................................................................................................... 16
      2.3.4.4. Formanten ....................................................................................................................... 16
    2.3.5. Verarbeitungsprozesse des Gehörs ......................................................................................... 16
      2.3.5.1. Art der Signalverarbeitung ............................................................................................... 17
      2.3.5.2. Wahrnehmungsgeschwindigkeit ..................................................................................... 18
  2.4. Kognitive und psychologische Grundlagen ............................................................................... 19
    2.4.1. Emotionen .............................................................................................................................. 19
      2.4.1.1. Einführung .......................................................................................................................... 19
      2.4.1.2. Emotion, Gefühl, Affekt, Stimmung und Empathie .......................................................... 20
      2.4.1.3. Definitionen des Begriffs „Emotion“ ............................................................................... 21
      2.4.1.4. Aspekte der Emotion ....................................................................................................... 22
      2.4.1.5. Emotionstheorien .......................................................................................................... 23
        2.4.1.5.1. Emotion nach Plutchik .............................................................................................. 24
      2.4.1.6. Emotionale Kommunikation ........................................................................................... 26
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.


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. Göttert3 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


2.1. Die Anfänge der Stimmforschung


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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{8} Der Mensch nimmt Tonhöhenunterschiede allerdings unterschiedlich stark wahr – so werden Unterschiede zwischen hochfrequenten Tönen als geringer eingestuft als solche zwischen tieffrequenten.\textsuperscript{9}

- **Amplitude:**
  
  Die Weite der Bewegung eines schwingenden Körpers, die die Stärke der Schallwellen angibt. Dieser Schalldruckpegel wird in Pascal (Pa) oder Mikropascal (\(\mu\)Pa) 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.\textsuperscript{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.\textsuperscript{11}

- **Klang:**
  
  Wird in der Alltagssprache dem Ton gleichgesetzt, bezeichnet allerdings richtigerweise

\textsuperscript{6} Vgl. Mathelitsch / Friedrich (1995), S. 5-8
\textsuperscript{7} Vgl. Zorowka / Höfler, in: Friedrich / Bigenzahn / Zorowka (2000), S. 337
\textsuperscript{8} Vgl. Mathelitsch / Friedrich (1995), S. 28 - 30
\textsuperscript{9} Vgl. Stassen (1995), S. 8
\textsuperscript{10} Vgl. Zorowka / Höfler, in: Friedrich / Bigenzahn / Zorowka (2000), S. 337
\textsuperscript{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.\textsuperscript{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.\textsuperscript{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

\textsuperscript{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.\(^\text{14}\) 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.\(^\text{15}\) Am empfindlichsten reagiert die Amygdala auf auditive Reize.\(^\text{16}\)

In der Großhirnrinde werden Sinnesindrücke 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 die 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 Reptilienhirn und dem Limbischen System nun verstanden und kontrolliert werden.\(^\text{17}\)


\(^\text{14}\) Vgl. Hülshoff (1999), S. 34 ff.
\(^\text{15}\) Vgl. Merten (2003), S. 86
\(^\text{16}\) Vgl. Ebenda, S. 94
\(^\text{17}\) 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.\textsuperscript{18}

Zum aktuellen Stand der Emotionsforschung im Bereich der Stimme siehe Kapitel 3.1.

\subsection*{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 zu steigern, 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.\textsuperscript{19}

\subsection*{2.3.2. Voraussetzungen für Stimmbildung und Stimmerkennung}

\subsubsection*{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.\textsuperscript{20}

\begin{flushright}
\end{flushright}
motorischen Sprachzentren – der Großhirnrinde und dem Frontallappen – erfährt die Reaktion auf das Erlebnis die Umwandlung in Sprache, Verbalisation genannt.  

- 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).  

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.

22 Vgl. Ebenda, S. 26
23 Vgl. Ebenda, S. 26
24 Vgl. Mathelitsch / Friedrich (1995), S. 46
2.3.2.2. Atmung und Sprechen


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.\(^\text{28}\)

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.

**2.3.2.3. Die Stimmproduktion**


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 Hörindrücke 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.\(^{38}\)

\(^{34}\) Vgl. Johnson-Laird (1996), S. 321
\(^{37}\) 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


\[^{42}\] Vgl. Ebenda, S. 54 f.
und eine vollständige Reifung zum Geburtstermin vorausgesetzt – schreien bei ihrer Geburt mit 400 bis 500 Hertz.\textsuperscript{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.\textsuperscript{45}

Vor der Pubertät liegt die Spanne des Grundtons zwischen 147 bis 698 Hz.\textsuperscript{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 Sprechstimmmlage 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.\textsuperscript{47} Im Erwachsenenalter liegt die mittlere Sprechstimmmlage beim Mann schließlich zwischen 100 Hz und 150 Hz, bei der Frau zwischen 200 Hz und 250 Hz.\textsuperscript{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.\textsuperscript{49} Die Stimme der Frau sinkt im Alter kontinuierlich, bei 80-bis 90-Jährigen liegt die mittlere Sprechstimmmlage etwa bei 195 Hz. Auch die Stimme des Mannes wird tiefer, allerdings nur bis etwa zum fünfzigsten Lebensjahr, woraufhin die Grundfrequenz allmählich steigt.\textsuperscript{50} Im Alter von 90 Jahren liegt die mittlere Sprechstimmmlage eines Mannes etwa zwischen 131 und 147 Hz.\textsuperscript{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

\textsuperscript{44} Vgl. Mathelitsch / Friedrich (1995), S. 55 f.  
\textsuperscript{45} Vgl. Gundermann (1994), S. 46  
\textsuperscript{46} Vgl. Ebenda, S. 46  
\textsuperscript{47} Vgl. Mathelitsch / Friedrich (1995), S. 61  
\textsuperscript{48} Vgl. Friedrich, in: Friedrich / Bigenzahn / Zorowka (2000), S. 56  
\textsuperscript{49} Vgl. Mathelitsch / Friedrich (1995), S. 68  
\textsuperscript{50} Vgl. Gundermann (1994), S. 46  
\textsuperscript{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.\textsuperscript{52}


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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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. Schwangerschaftwoche ist – als erstes Sinnesorgan – die Schnecke und damit die körperliche Voraussetzung für das Gehör voll ausgebildet\textsuperscript{57}, etwa zwei Wochen später beginnen Föten schließlich zu hören. Schon zehn Wochen später haben Ungeborene ein funktionales, auf andere bezogenes Gehör.\textsuperscript{58} Bei Föten

\textsuperscript{53} Vgl. Ebenda, S. 226
\textsuperscript{54} Vgl. Ebenda, S. 226 ff.
\textsuperscript{55} Vgl. Birbaumer / Schmidt, in: Schmidt / Schaible (Hg.) (2006), S. 450
\textsuperscript{56} Vgl. Ebenda, S. 454
\textsuperscript{57} Vgl. Wickel, / Hartogh (2006), S. 40
\textsuperscript{58} Vgl. Chamberlain (1990), S. 55
sind laut Dr. Tomatis\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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

\textsuperscript{59} Vgl. Chamberlain (1990), S. 57
\textsuperscript{60} Vgl. Wickel / Hartogh (2006), S. 43
\textsuperscript{61} Vgl. Ebenda, S. 46 f.
\textsuperscript{62} Vgl. Stelzig, in: Lotzmann (Hg.) (1997), S. 9
\textsuperscript{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.\textsuperscript{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).\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{68}

### 2.3.4.2. Stimmgattungen


\textsuperscript{65} Vgl. Ebenda, S. 56
\textsuperscript{66} Vgl. Siegmüller / Bartels (2006), S. 362
\textsuperscript{67} Vgl. Friedrich, in: Friedrich / Bigenzahn / Zorowka (2000), S. 56
\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.


\begin{footnotes}
\item[74] Vgl. Chion (1994), S. 10 f.
\item[75] Vgl. Goldstein / Irtel (Hg.) (2008), S. 123
\item[76] Vgl. Chion (1994), S. 12 f.
\item[77] Vgl. Gruhn (2005), S. 13
\end{footnotes}
verarbeitet und somit aktiv wahrgenommen werden, die anhand der Schallmuster als wichtig eingestuft wurden.\footnote{Vgl. Zenner, in: Schmidt / Schaible (Hg.) (2006), S. 308}

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 Herkunftsart verändert.\footnote{Vgl. Römer (1994), S. 40 f.}

Ähnlich den Regeln der Gestaltpsychologie (siehe 2.8.3) gruppiert das auditorische System die ankommenden Reize, was die Verarbeitung vereinfacht. So werden Schalleignisse 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.\footnote{Vgl. Goldstein / Irtel (Hg.) (2008), S. 302 f.}

\subsection*{2.3.5.2. Wahrnehmungsgeschwindigkeit}


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

\footnote{Vgl. Zenner, in: Schmidt / Schaible (Hg.) (2006), S. 308}
\footnote{Vgl. Römer (1994), S. 40 f.}
\footnote{Vgl. Goldstein / Irtel (Hg.) (2008), S. 302 f.}
\footnote{Vgl. Sonnenschein (2001), S. 90}
\footnote{Vgl. Grau, in: Felsmann (Hg.) (2008), S. 39}
0,5 bis 1 Sekunde. 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. Ab einer Länge von 50 Millisekunden wird die Lautstärke wahrgenommen, für die Klangfarbe eines Geräusches sowie die Erkennung von Konsonanten in der Sprache ist 0,1 Sekunde nötig. 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.

2.4. Kognitive und psychologische Grundlagen

2.4.1. Emotionen

2.4.1.1. Einführung

Hülshoff definiert Emotionen als „körperlich-seelische Reaktionen, durch die ein Umweltreignis 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.

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. 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.


2.4.1.2. Emotion, Gefühl, Affekt, Stimmung und Empathie


Stassen 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

„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.\textsuperscript{96}


**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.“\textsuperscript{97}

\textsuperscript{94} Vgl. Merten (2003), S. 11
\textsuperscript{95} Zitiert nach Smith, in: Bartsch / Eder / Fahlenbrach (Hg.) (2007), S. 43
\textsuperscript{96} Vgl. Tischer (1993), S. 5
\textsuperscript{97} Zitiert nach Merten (2003), S. 355
Ebenfalls weite Verbreitung hat die Definition von Oatley und Jenkins gefunden, die Handlungsbereitschaft in den Mittelpunkt rückt:


2.4.1.4. Aspekte der Emotionen


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 kategorischem Ansatz. Emotionen können laut dimensionalem Ansatz auf Basis weniger, grundlegenden 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.\(^{103}\)


\(^{100}\) 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


„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.\(^{105}\)


„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 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“


107 Vgl. Plutchik (2003), S. 105 f.
108 Vgl. Ebenda, S. 109
2.4.1.6. Emotionale Kommunikation


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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{114}

\textsuperscript{110} Vgl. Merten (2003), S. 144 ff.
\textsuperscript{111} Vgl. Ebenda, S. 148
\textsuperscript{112} Vgl. Ebenda, S. 148
\textsuperscript{113} Vgl. Ebenda, S. 156
\textsuperscript{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 Teilspekte 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

2.4.1.8. Erlernen des Emotionslexikons


\textsuperscript{119} Vgl. Tischer (1993), S. 53 f.
\textsuperscript{120} Vgl. Battacchi (1997), S. 74 ff.
Verwendung der einzelnen Emotionswörter ist von den sozialen und kulturellen Eigenheiten der Muttersprache und des Landes abhängig.  

2.4.2. Stimme, Sprache und Gehör im kognitiven Entwicklungsprozess

2.4.2.1. Der Fötus – Stimme im Mutterleib


Chamberlain geht aufgrund der Erkenntnisse zum Hörbereich von Föten davon aus, dass das Ungeborene die hohe Stimme der Mutter besser hört als die tiefe 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 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.

122 Vgl. Chamberlain (1990), S. 58
123 Vgl. Ebenda, S. 56 f.
125 Vgl. Chamberlain (1990), S. 57
unterscheiden können. Dieser Meinung schließt sich auch Westphal\textsuperscript{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“.

\subsection*{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.\textsuperscript{128} Das Gehör gesunder Neugeborener ist bereits so gut wie das Erwachsener, allerdings nehmen Neugeborene Laute Chamberlain\textsuperscript{129} zulasten 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.\textsuperscript{130} Von der Geburt bis etwa zum ersten Lebensjahr können Babys Phoneme, also die kleinsten Spracheinheiten (etwa „ba“ und „pa“) besser unterscheiden als Erwachsene.\textsuperscript{131} Bereits unmittelbar nach der Geburt kann ein Kind außerdem verschiedene Sprachen erkennen, es differenziert anhand verschiedenster, auch nicht-muttersprachlicher Lautkontraste.\textsuperscript{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 Stimmhäuter schont.\textsuperscript{133} Ab der 4. bis 5. Lebenswoche kann das Kind Tonhöhe, Lautstärke, Stimmeinsatz sowie Tonhöhen- und Lautstärkenverlauf des Schreisens bewusst verändern, der „modulierte Schrei“ tritt damit erstmals auf.\textsuperscript{134}

\begin{flushright}
\textsuperscript{127} Vgl. Westphal (2002), S. 105
\textsuperscript{128} Vgl. Denk / Brunner / Bigenzahn, in: Friedrich / Bigenzahn / Zorowka (2000), S. 244
\textsuperscript{129} Vgl. Chamberlain (1990), S. 60
\textsuperscript{130} Vgl. Clark / Clark (1977), S. 376
\textsuperscript{131} Vgl. Chamberlain (1990), S. 122
\textsuperscript{133} Vgl. Papoušek / Papoušek, in: Scherer (1982), S. 82
\textsuperscript{134} Vgl. Mathelitsch / Friedrich (1995), S. 57
\end{flushright}
2.4.2.3. Gurren und Lallen

Von etwa der siebenten Woche an bis zum 3. Lebensmonat erweitern Gurrlaute das Repertoire des Kleinkindes. 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. 


2.4.2.4. Sprachverständnis und erste Worte


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

---

136 Vgl. Clark / Clark (1977), S. 377
138 Vgl. Clark / Clark (1977), S. 377
142 Vgl. Clark / Clark (1977), S. 389


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.

Sonnenschein 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 herausfiltern.

Vgl. Ebenda, S. 149
Vgl. Ebenda, S. 149
Vgl. Sonnenschein (2001), S. 137


\subsection{2.4.3.1. Arten des Zuhörens}

Chion\textsuperscript{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)


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

Sonnenschein\textsuperscript{152} schlägt eine weitere Art des Zuhörrens als Unterscheidung vor: „Referential Listening“ meint, dass der Zuhörer sich der Bedeutung eines Geräusches 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.

\subsection*{2.4.3.2. Kommunikation}

Geißner\textsuperscript{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“\textsuperscript{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 ererbte soziale Ausdrucksmuster.“\textsuperscript{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.\textsuperscript{156}

\subsection*{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

\begin{itemize}
\item[\textsuperscript{151}] Vgl. Chion (1994), S. 28 ff.
\item[\textsuperscript{152}] Vgl. Sonnenschein (2001), S. 77 f.
\item[\textsuperscript{153}] Vgl. Geißner (1981), S. 81
\item[\textsuperscript{154}] Vgl. Ebenda, S. 96
\item[\textsuperscript{155}] Vgl. Ebenda, S. 120
\item[\textsuperscript{156}] Vgl. Geißner (1981), S. 123
\end{itemize}

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.\textsuperscript{158}

\subsection*{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.\textsuperscript{159}

Für das Erkennen eines uns bekannten Menschen genügt bereits ein Husten oder sogar eine charakteristische Art und Weise, Luft zu holen, wie Sonnenschein\textsuperscript{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.\textsuperscript{161}

\begin{footnotesize}
\begin{enumerate}
\item[158] Vgl. Ebenda, S. 165
\item[159] Vgl. Chion (1994), S. 27
\item[160] Vgl. Sonnenschein (2001), S. 132
\item[161] Vgl. Habermann (2001), S. 86
\end{enumerate}
\end{footnotesize}
Sich eine Stimme in Erinnerung zu rufen, fällt allerdings schwer – Chion\textsuperscript{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“\textsuperscript{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.\textsuperscript{164}

2.4.4.3. Stimme sticht hervor
Sonnenschein\textsuperscript{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.\textsuperscript{166}

\textsuperscript{162} Vgl. Chion (1999), S. 1  
\textsuperscript{163} Vgl. Ebenda, S.5  
\textsuperscript{164} Vgl. Ebenda, S.5  
\textsuperscript{165} Vgl. Sonnenschein (2001), S. 80  
\textsuperscript{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.
\(^{169}\) Vgl. Stelzig, in: Lotzmann (1997), S. 6
Transformation vom visuellen zum optischen Reiz. Postman\textsuperscript{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.\textsuperscript{171}

2.5.3. Die Rolle der Stimme in der Sprache

Die Stimme hat laut Stelzig\textsuperscript{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 unmittelbare 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.\textsuperscript{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 Angstbemildung, Hemmsignale zum Verscheuchen oder Rufsignale zur Annäherung an andere sowie Sanftheit zur Vertrauensbildung.\textsuperscript{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

\textsuperscript{170} Vgl. Postman (2006), S. 43
\textsuperscript{171} Vgl. Stelzig, in: Lotzmann (1997), S. 7
\textsuperscript{172} Vgl. Ebenda, S. 1
\textsuperscript{173} Vgl. Ebenda, S. 4
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.\(^\text{175}\)

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.\(^\text{176}\)


**2.6.1. Enkodierung von Emotionen durch die Stimme**

Johnson-Laird\(^\text{178}\) 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  
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\textsuperscript{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\textsuperscript{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\textsuperscript{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

\textsuperscript{179} Vgl. Tischer (1993), S. 17
\textsuperscript{180} Vgl. Williams / Stevens, in: Scherer (1982), S. 308
\textsuperscript{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.


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
\(^ {185}\) Vgl. Ebenda, S. 301
\(^ {186}\) Vgl. Scherer, in Scherer (1982), S. 300

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.\textsuperscript{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 zwangsläufig ein Lügner sein – diese Fehlinterpretation wird als „Othello-Fehler“ bezeichnet.\textsuperscript{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.\textsuperscript{189}

2.6.3. Dekodierung von Emotionen in der Stimme
Laut Chamberlain\textsuperscript{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.

\textsuperscript{187} Vgl. Merten (2003), S. 139  
\textsuperscript{188} Vgl. Ekman (1989), S. 70 f.  
\textsuperscript{189} Vgl. Gundermann (1994), S. 35  
\textsuperscript{190} Vgl. Chamberlain (1990), S. 60 f.

Scherer\textsuperscript{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\textsuperscript{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\textsuperscript{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

\textsuperscript{191} Vgl. Chamberlain (1990), S. 115 f.
\textsuperscript{192} Vgl. Scherer, in Scherer (1982), S. 303
\textsuperscript{193} Vgl. Williams / Stevens, in Scherer (1982), S. 313 ff.
\textsuperscript{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


2.7.2. Anfänge des Fernsehens in Österreich


196 Vgl. Faulstich (2005), S. 55


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


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 Programmspots. 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 dieser Platz mit 41,4 Prozent der Zeit die Fernsehpublizistik (dazu zählen z.B. Nachrichtensendungen, Magazine, Reportagen sowie Talkformate und Dokumentationen) ein.


---

199 Vgl. Ebenda, S. 27 f.
200 Vgl. Steininger / Woelke (Hg.) (2008): S. 32
2.7.4. Fernsehserien

Die typische Fernsehserie zeichnet sich durch ihren fiktionalen Charakter aus. Faulstich\textsuperscript{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.\textsuperscript{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).\textsuperscript{205}

2.7.4.1. Krimiserien


\textsuperscript{203} Vgl. Faulstich (2008), S 106
\textsuperscript{204} Vgl Ebenda, S. 108 f.
Unterkategorien herausgebildet, etwa Detektiv-, Polizei-, Gerichts- oder Spionagefilm sowie Thriller.\footnote{Vgl. Hickethier (2005), S. 14 ff.}

Ein wahrer Boom an fiktionalen Krimiserien im Fernsehen setzte im deutsdsprachigen Raum in den 1970er Jahren mit „Derrik“, „Der Kommissar“, „Tatort“ und anderen ein.\footnote{Vgl Faulstich (2008), S. 109}

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


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 oft mal zu Krisen.\footnote{Vgl. Bignell (2008), S. 130}

### 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.\footnote{Vgl. Schlimbach (2007), S. 194}

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

\begin{flushright}
\footnotesize
\textsuperscript{206} Vgl. Hickethier (2005), S. 14 ff.  
\textsuperscript{207} Vgl Faulstich (2008), S. 109  
\textsuperscript{208} Vgl. Burton (2000), S. 209  
\textsuperscript{209} Vgl. Bignell (2008), S. 130  
\textsuperscript{210} Vgl. Schlimbach (2007), S. 194
\end{flushright}


Um den Inhalt eines fiktiven Films oder Fernsehinhalts zu verstehen, sind laut Ohler 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 spricht von der Kombination als „narratives Form-Inhalt-Korrespondenzgitter“. Gewisse

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.\textsuperscript{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.\textsuperscript{215} Ebenso umstritten wie die Art des Hineinlebens sind die konkreten filmischen Werkzeuge, mit denen Empathie beim Zuschauer ausgelöst wird.

Mikunda\textsuperscript{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

\textsuperscript{216} Vgl. Mikunda (2002), S. 34 ff.

2.7.7. Spezifika des Fernsehens


Fernsehen ist weniger von Stars per Kino-Definition abhängig sondern mehr von Persönlichkeiten, denen von den Zuschauern bestimmte Qualitäten zugeschrieben werden.

Durch die nur gelegentliche Berührung mit dem Kino entsteht zu dessen Stars laut Burton\textsuperscript{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


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.


\textsuperscript{219} Vgl. Burton (2000), S 77 f.  
\textsuperscript{220} Vgl. Westphal (2002), S.157
normalerweise von jenem Ort stammt, an dem wir eine zugehörige Schallquelle sehen.\textsuperscript{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.\textsuperscript{222}

Chion\textsuperscript{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.


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.\textsuperscript{225}

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

2.8.1.1. Zusatznutzen
Chion\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{231} Besondere Bedeutung hat der Zusatznutzen des Tons Chion zufolge dank des „Synchresis“-Prinzips.

\textsuperscript{227} Vgl. Ebenda, S. 48 ff.
\textsuperscript{228} Vgl. Ebenda, S. 5
\textsuperscript{229} Vgl. Ebenda, S. 7
\textsuperscript{230} Vgl. Chion (1994), S. 10 f.
\textsuperscript{231} Vgl. Ebenda, S. 22 f.
2.8.1.2. „Synchresis“ und Synchronizität


2.8.2. Unterteilungskriterien

Man kann Sound im audiovisuellen Medium nach verschiedenen Kriterien einteilen – Zettl 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.

---

233 Vgl. Chion (1994), S. 63
234 Vgl. Ebenda, S. 58 ff.


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.\textsuperscript{239}

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


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

\textsuperscript{240} Vgl. Sonnenschein (2001), S. 79

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. 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 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.

2.9. Stimme im audiovisuellen Medium

Chion\textsuperscript{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.\textsuperscript{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\textsuperscript{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 abgehobelt existiert.

Nicht nur technisch unterscheidet sich Stimme im Medium vom üblichen natürlichen Höreindruck. Sonnenschein\textsuperscript{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.”.\textsuperscript{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

\begin{itemize}
  \item \textsuperscript{243} Vgl. Chion (1994), S. 5 f.
  \item \textsuperscript{244} Vgl Bignell (2008), S. 157
  \item \textsuperscript{245} Vgl. Westphal (2002), S. 157
  \item \textsuperscript{246} Vgl. Sonnenschein (2001), S. 146
  \item \textsuperscript{247} Vgl. Ebenda, S. 146 f.
\end{itemize}
der üblichen Frage bei einem Handygespräch: „Wo bist du?“. 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

2.9.2. Geschichte der deutschsprachigen Synchronisation

249 Vgl. Ebenda, S. 156 f.
250 Vgl. Westphal (2002), S. 158

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.


\textsuperscript{252} Vgl. Bräutigam (2001), S. 10 ff.
\textsuperscript{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. 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.

Durch die Synchronstimme werden dem Schauspieler bestimmte Persönlichkeitsmerkmale zugeordnet. Bräutigam zufolge ist daher weniger wichtig, ob die Originalstimme der Synchronstimme ähnelt, als wie die Synchronstimme zum Charakter in Film oder Serie passt. Gundermann 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.

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
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.\textsuperscript{259}


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\textsuperscript{262} zufolge ist dies vor allem deshalb nötig, weil im audiovisuellen Medium jene stereophonan 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\textsuperscript{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.

\textsuperscript{259} Vgl. Stassen (1995), S. 8
\textsuperscript{260} Vgl. Sonnenschein (2001), S. 32 f.
\textsuperscript{261} Vgl. Ebenda, S. 32 f.
\textsuperscript{262} Vgl. Chion (1999), S.5
\textsuperscript{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 festhält. Anhand verschiedener Untersuchungen ist zwar inzwischen zweifelsfrei geklärt, dass Emotionen über die Stimme kodiert und von Zuhöern 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.


---

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

Bei einer Untersuchung 2007 verglichen Kreifelts et al.\textsuperscript{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.\textsuperscript{268} im Jahr 2005, da auch sie der Ansicht waren, dass sich lediglich wenige Studien mit der

\textsuperscript{266} Vgl. Wildgruber et al., in: NeuroImage (2005), S. 1233 ff.


### 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 4.2: 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


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.\textsuperscript{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

\textsuperscript{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**

**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 von 137 Items misst. Er dient sowohl zum Vergleich großer Personengruppen als auch als individueller psychologischer Test. Das Freiburger


**Vorteile**


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**

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 \((r = .20 \text{ bis } r = .30)\) 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**


\(^{275}\) 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.


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**

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 Befragungs teilnehmern 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


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 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 bekamen 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 erklängt) vertraut ist – wenn auch unterbewusst.


\[279\] 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


Die sieben kurzen Ausschnitte stammen aus diesen Folgen:

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“:


„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

„CSI: Miami“

„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 verbinden einerseits, dass Horatios Bruder mit Yelina verheiratet war, bevor er als krimineller Polizist enttarnt und getötet wurde. Außerdem entdecken die beiden, dass sie Gefühle für einander 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 für einander 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 vierte Teil war der auditive, visuelle und audiovisuelle Fragebogenteil. Dabei standen am Anfang einige Instruktionen, die zum Beispiel darauf hinwiesen, dass die Audio-/Video-


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

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

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 nachzuverziehen, 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.

<table>
<thead>
<tr>
<th>Altersgruppen</th>
<th>Häufigkeit</th>
<th>Prozent</th>
<th>Gültige Prozent</th>
<th>Kumulierte Prozent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gültig bis 19 Jahre</td>
<td>2</td>
<td>1,9</td>
<td>1,9</td>
<td>1,9</td>
</tr>
<tr>
<td>20 bis 29 Jahre</td>
<td>43</td>
<td>41,3</td>
<td>41,7</td>
<td>43,7</td>
</tr>
<tr>
<td>30 bis 39 Jahre</td>
<td>15</td>
<td>14,4</td>
<td>14,6</td>
<td>58,3</td>
</tr>
<tr>
<td>40 bis 49 Jahre</td>
<td>10</td>
<td>9,6</td>
<td>9,7</td>
<td>68,0</td>
</tr>
<tr>
<td>50 bis 59 Jahre</td>
<td>18</td>
<td>17,3</td>
<td>17,5</td>
<td>85,4</td>
</tr>
<tr>
<td>60 bis 69 Jahre</td>
<td>14</td>
<td>13,5</td>
<td>13,6</td>
<td>99,0</td>
</tr>
<tr>
<td>ab 70 Jahre</td>
<td>1</td>
<td>1,0</td>
<td>1,0</td>
<td>100,0</td>
</tr>
<tr>
<td>Gesamt</td>
<td>103</td>
<td>99,0</td>
<td>100,0</td>
<td></td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Beruf</th>
<th>Häufigkeit</th>
<th>Prozent</th>
<th>Gültige Prozent</th>
<th>Kumulierte Prozent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fehlend</td>
<td>1</td>
<td>1,0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gesamt</td>
<td>104</td>
<td>100,0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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.

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.
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.

<table>
<thead>
<tr>
<th></th>
<th>Häufigkeit</th>
<th>Prozent</th>
<th>Gültige Prozente</th>
<th>Kumulierte Prozente</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gültig</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ja</td>
<td>43</td>
<td>41,3</td>
<td>41,3</td>
<td>41,3</td>
</tr>
<tr>
<td>nein</td>
<td>61</td>
<td>58,7</td>
<td>58,7</td>
<td>100,0</td>
</tr>
<tr>
<td>Gesamt</td>
<td>104</td>
<td>100,0</td>
<td>100,0</td>
<td></td>
</tr>
</tbody>
</table>

Altersgruppen * Kinder Kreuztabelle

<table>
<thead>
<tr>
<th>Altersgruppen</th>
<th>Kinder</th>
<th>Gesamt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ja</td>
<td>nein</td>
</tr>
<tr>
<td>bis 19 Jahre</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>% innerhalb von Altersgruppen</td>
<td>0,0%</td>
<td>100,0%</td>
</tr>
<tr>
<td>20 bis 29 Jahre</td>
<td>0</td>
<td>43</td>
</tr>
<tr>
<td>% innerhalb von Altersgruppen</td>
<td>0,0%</td>
<td>100,0%</td>
</tr>
<tr>
<td>30 bis 39 Jahre</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>% innerhalb von Altersgruppen</td>
<td>20,0%</td>
<td>80,0%</td>
</tr>
<tr>
<td>40 bis 49 Jahre</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>% innerhalb von Altersgruppen</td>
<td>80,0%</td>
<td>20,0%</td>
</tr>
<tr>
<td>50 bis 59 Jahre</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>% innerhalb von Altersgruppen</td>
<td>88,9%</td>
<td>11,1%</td>
</tr>
<tr>
<td>60 bis 69 Jahre</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>% innerhalb von Altersgruppen</td>
<td>100,0%</td>
<td>0,0%</td>
</tr>
<tr>
<td>ab 70 Jahre</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>% innerhalb von Altersgruppen</td>
<td>100,0%</td>
<td>0,0%</td>
</tr>
<tr>
<td>Gesamt</td>
<td>42</td>
<td>61</td>
</tr>
<tr>
<td>% innerhalb von Altersgruppen</td>
<td>40,8%</td>
<td>59,2%</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mittelwert</th>
<th>Standardabweichung</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alle Videos - Auditiv richtig erkannt in Prozent</td>
<td>104</td>
<td>39,5032</td>
<td>25,01335</td>
<td>0,00</td>
<td>100,00</td>
</tr>
<tr>
<td>Auditiv richtig erkannt - Erwartungswert</td>
<td>104</td>
<td>23,7308</td>
<td>1,87073</td>
<td>21,00</td>
<td>25,00</td>
</tr>
</tbody>
</table>

### Ränge

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mittlerer Rang</th>
<th>Rangsumme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auditiv richtig erkannt - Negative Ränge</td>
<td>64*</td>
<td>50,13</td>
<td>3208,00</td>
</tr>
<tr>
<td>Erwartungswert - Alle Videos - Positive Ränge</td>
<td>24b</td>
<td>29,50</td>
<td>708,00</td>
</tr>
<tr>
<td>Auditiv richtig erkannt in Prozent - Bindungen</td>
<td>16c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gesamt</td>
<td>104</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Z</th>
<th>Asymptotische Signifikanz (2-seitig)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-5,252</td>
<td>0,000</td>
</tr>
</tbody>
</table>

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 konnten. 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

<table>
<thead>
<tr>
<th>Häufigkeit</th>
<th>Prozent</th>
<th>Gültige Prozente</th>
<th>Kumulierte Prozente</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gültig</td>
<td>12</td>
<td>11,5</td>
<td>11,5</td>
</tr>
<tr>
<td>16,67</td>
<td>12</td>
<td>11,5</td>
<td>23,1</td>
</tr>
<tr>
<td>25,00</td>
<td>16</td>
<td>15,4</td>
<td>38,5</td>
</tr>
<tr>
<td>33,33</td>
<td>19</td>
<td>18,3</td>
<td>56,7</td>
</tr>
<tr>
<td>50,00</td>
<td>16</td>
<td>15,4</td>
<td>72,1</td>
</tr>
<tr>
<td>66,67</td>
<td>24</td>
<td>23,1</td>
<td>95,2</td>
</tr>
<tr>
<td>75,00</td>
<td>1</td>
<td>1,0</td>
<td>96,2</td>
</tr>
<tr>
<td>Häufigkeit</td>
<td>Prozent</td>
<td>Gültige Prozente</td>
<td>Kumulierte Prozente</td>
</tr>
<tr>
<td>------------</td>
<td>---------</td>
<td>------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Gültig</td>
<td>12</td>
<td>11,5</td>
<td>11,5</td>
</tr>
<tr>
<td>33,33</td>
<td>12</td>
<td>11,5</td>
<td>23,1</td>
</tr>
<tr>
<td>50,00</td>
<td>21</td>
<td>20,2</td>
<td>43,3</td>
</tr>
<tr>
<td>66,67</td>
<td>14</td>
<td>13,5</td>
<td>56,7</td>
</tr>
<tr>
<td>100,00</td>
<td>40</td>
<td>38,5</td>
<td>95,2</td>
</tr>
<tr>
<td>150,00</td>
<td>5</td>
<td>4,8</td>
<td>100,0</td>
</tr>
<tr>
<td>Gesamt</td>
<td>104</td>
<td>100,0</td>
<td>100,0</td>
</tr>
</tbody>
</table>

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“

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:

<table>
<thead>
<tr>
<th>Statistiken</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Emotion pro Sequenz auditiv</td>
</tr>
<tr>
<td>richtig erkannt in Prozent</td>
</tr>
<tr>
<td>N Gültig</td>
</tr>
<tr>
<td>Fehlend</td>
</tr>
<tr>
<td>Mittelwert</td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Statistiken</th>
<th>Traurigkeit auditiv richtig erkannt in Prozent</th>
<th>Überraschung auditiv richtig erkannt in Prozent</th>
<th>Vertrauen auditiv richtig erkannt in Prozent</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Gültig</td>
<td>71</td>
<td>104</td>
<td>71</td>
</tr>
<tr>
<td>Fehlend</td>
<td>33</td>
<td>0</td>
<td>33</td>
</tr>
<tr>
<td>Mittelwert</td>
<td>26,7606</td>
<td>25,9615</td>
<td>28,1690</td>
</tr>
<tr>
<td>Modus</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Standardabweichung</td>
<td>44,58618</td>
<td>44,05467</td>
<td>45,30247</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistiken</th>
<th>Angst auditiv richtig erkannt in Prozent</th>
<th>Wut auditiv richtig erkannt in Prozent</th>
<th>Neugierde auditiv richtig erkannt in Prozent</th>
<th>Freude auditiv richtig erkannt in Prozent</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Gültig</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>35</td>
</tr>
<tr>
<td>Fehlend</td>
<td>71</td>
<td>71</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Mittelwert</td>
<td>93,9394</td>
<td>39,3939</td>
<td>49,61977</td>
<td>.0000</td>
</tr>
<tr>
<td>Modus</td>
<td>100,00</td>
<td>100,00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Standardabweichung</td>
<td>24,23058</td>
<td>50,63094</td>
<td>49,61977</td>
<td>.000000</td>
</tr>
</tbody>
</table>

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.

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 reiniger 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](image)
a. Die zu testende Verteilung ist eine Normalverteilung.
b. Aus den Daten berechnet.

<table>
<thead>
<tr>
<th>Ränge</th>
<th>N</th>
<th>Mittlerer Rang</th>
<th>Rangsumme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative Ränge</td>
<td>35(^b)</td>
<td>47,49</td>
<td>1662,00</td>
</tr>
<tr>
<td>Positive Ränge</td>
<td>64(^b)</td>
<td>51,38</td>
<td>3288,00</td>
</tr>
<tr>
<td>Bindungen</td>
<td>5(^c)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gesamt</td>
<td>104</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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\(^b\)**

<table>
<thead>
<tr>
<th>Alle Videos - Visuell richtig erkannt in Prozent - Alle Videos - Auditiv richtig erkannt in Prozent</th>
<th>Z</th>
<th>Asymptotische Signifikanz (2-seitig)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-2,846(^d)</td>
<td>.004</td>
</tr>
</tbody>
</table>

a. Basiert auf negativen Rängen.
b. Wilcoxon-Test

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

**Statistiken**

<table>
<thead>
<tr>
<th>Alle Videos - Auditiv richtig erkannt - Gruppen Prozent</th>
<th>Alle Videos - Visuell richtig erkannt - Gruppen Prozent</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Gültig</td>
<td>104</td>
</tr>
<tr>
<td>Fehlend</td>
<td>0</td>
</tr>
</tbody>
</table>

Alle Videos - Auditiv richtig erkannt - Gruppen Prozent
<table>
<thead>
<tr>
<th>Häufigkeit</th>
<th>Prozent</th>
<th>Gültige Prozente</th>
<th>Kumulierte Prozente</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gültig</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 % richtig erkannt</td>
<td>12</td>
<td>11,5</td>
<td>11,5</td>
</tr>
<tr>
<td>0,1 bis 25 % richtig erkannt</td>
<td>28</td>
<td>26,9</td>
<td>38,5</td>
</tr>
<tr>
<td>25,1 bis 50 % richtig erkannt</td>
<td>35</td>
<td>33,7</td>
<td>72,1</td>
</tr>
<tr>
<td>50,1 bis 75 % richtig erkannt</td>
<td>25</td>
<td>24,0</td>
<td>96,2</td>
</tr>
<tr>
<td>mehr als 75 % richtig erkannt</td>
<td>4</td>
<td>3,8</td>
<td>100,0</td>
</tr>
<tr>
<td>Gesamt</td>
<td>104</td>
<td>100,0</td>
<td>100,0</td>
</tr>
</tbody>
</table>

Alle Videos - Visuell richtig erkannt - Gruppen Prozent

<table>
<thead>
<tr>
<th>Häufigkeit</th>
<th>Prozent</th>
<th>Gültige Prozente</th>
<th>Kumulierte Prozente</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gültig</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 % richtig erkannt</td>
<td>2</td>
<td>1,9</td>
<td>1,9</td>
</tr>
<tr>
<td>0,1 bis 25 % richtig erkannt</td>
<td>14</td>
<td>13,5</td>
<td>15,4</td>
</tr>
<tr>
<td>25,1 bis 50 % richtig erkannt</td>
<td>63</td>
<td>60,6</td>
<td>76,0</td>
</tr>
<tr>
<td>50,1 bis 75 % richtig erkannt</td>
<td>22</td>
<td>21,2</td>
<td>97,1</td>
</tr>
<tr>
<td>mehr als 75 % richtig erkannt</td>
<td>3</td>
<td>2,9</td>
<td>100,0</td>
</tr>
<tr>
<td>Gesamt</td>
<td>104</td>
<td>100,0</td>
<td>100,0</td>
</tr>
</tbody>
</table>
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 erkannten, 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.

<table>
<thead>
<tr>
<th>Kolmogorov-Smirnov-Anpassungstest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traurigkeit auditiv richtig erkannt in Prozent</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>Parameter der Normalverteilung a, b</td>
</tr>
<tr>
<td>Mittelwert</td>
</tr>
<tr>
<td>Standardabweichung</td>
</tr>
<tr>
<td>Extremste Differenzen</td>
</tr>
<tr>
<td>Absolut</td>
</tr>
<tr>
<td>Positiv</td>
</tr>
<tr>
<td>Negativ</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov-Z</td>
</tr>
<tr>
<td>Asymptotische Signifikanz (2-seitig)</td>
</tr>
</tbody>
</table>

a. Die zu testende Verteilung ist eine Normalverteilung.
b. Aus den Daten berechnet.

<table>
<thead>
<tr>
<th>Kolmogorov-Smirnov-Anpassungstest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vertrauen auditiv richtig erkannt in Prozent</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>Parameter der Normalverteilung a, b</td>
</tr>
<tr>
<td>Mittelwert</td>
</tr>
<tr>
<td>Standardabweichung</td>
</tr>
<tr>
<td>Extremste Differenzen</td>
</tr>
<tr>
<td>Absolut</td>
</tr>
<tr>
<td>Positiv</td>
</tr>
<tr>
<td>Negativ</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov-Z</td>
</tr>
<tr>
<td>Asymptotische Signifikanz (2-seitig)</td>
</tr>
</tbody>
</table>

a. Die zu testende Verteilung ist eine Normalverteilung.
b. Aus den Daten berechnet.
Traurigkeit:

Ränge

<table>
<thead>
<tr>
<th>Rangtype</th>
<th>N</th>
<th>Mittlerer Rang</th>
<th>Rangsumme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative Ränge</td>
<td>1a</td>
<td>13,00</td>
<td>13,00</td>
</tr>
<tr>
<td>Positive Ränge</td>
<td>24b</td>
<td>13,00</td>
<td>312,00</td>
</tr>
<tr>
<td>Bindungen</td>
<td>11c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gesamt</td>
<td>36</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 Testb

<table>
<thead>
<tr>
<th></th>
<th>Traurigkeit visuell richtig erkannt in Prozent - Traurigkeit auditiv richtig erkannt in Prozent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z</td>
<td>-4,600a</td>
</tr>
<tr>
<td>Asymptotische Signifikanz (2-seitig)</td>
<td>.000</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>Traurigkeit auditiv richtig erkannt in Prozent</th>
<th>Traurigkeit visuell richtig erkannt in Prozent</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Gültig</td>
<td>71</td>
<td>69</td>
</tr>
<tr>
<td>Fehlend</td>
<td>33</td>
<td>35</td>
</tr>
</tbody>
</table>
Vertrauen:

<table>
<thead>
<tr>
<th>Ränge</th>
<th>N</th>
<th>Mittlerer Rang</th>
<th>Rangsumme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vertrauen visuell richtig erkannt in Prozent - Vertrauen auditiv richtig erkannt in Prozent</td>
<td>3a</td>
<td>4,50</td>
<td>13,50</td>
</tr>
<tr>
<td>Positive Ränge</td>
<td>5b</td>
<td>4,50</td>
<td>22,50</td>
</tr>
<tr>
<td>Bindungen</td>
<td>28c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gesamt</td>
<td>36</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 Test²

<table>
<thead>
<tr>
<th>Vertrauen visuell richtig erkannt in Prozent - Vertrauen auditiv richtig erkannt in Prozent</th>
<th>Z</th>
<th>Asymptotische Signifikanz (2-seitig)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-.707²</td>
<td>.480</td>
</tr>
</tbody>
</table>

a. Basiert auf negativen Rängen.
b. Wilcoxon-Test

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

<table>
<thead>
<tr>
<th></th>
<th>Gültig</th>
<th>Fehlend</th>
<th>Gültig</th>
<th>Fehlend</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>71</td>
<td>33</td>
<td>69</td>
<td>35</td>
</tr>
<tr>
<td>Mittelwert</td>
<td>28,1690</td>
<td>44,9275</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modus</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
</tbody>
</table>

**Ekel:**

<table>
<thead>
<tr>
<th></th>
<th>Ekel auditiv richtig erkannt in Prozent</th>
<th>Ekel visuell richtig erkannt in Prozent</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Gültig</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Fehlend</td>
<td>68</td>
</tr>
<tr>
<td>Mittelwert</td>
<td>37,500</td>
<td>5,8824</td>
</tr>
<tr>
<td>Modus</td>
<td>50,00</td>
<td>.00</td>
</tr>
<tr>
<td>Standardabweichung</td>
<td>27,71024</td>
<td>23,70435</td>
</tr>
</tbody>
</table>

**Ekel auditiv richtig erkannt in Prozent**

<table>
<thead>
<tr>
<th></th>
<th>Häufigkeit</th>
<th>Prozent</th>
<th>Gültige Prozente</th>
<th>Kumulierte Prozente</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gültig</td>
<td>0,00</td>
<td>11</td>
<td>10,6</td>
<td>30,6</td>
</tr>
<tr>
<td></td>
<td>50,00</td>
<td>23</td>
<td>22,1</td>
<td>63,9</td>
</tr>
<tr>
<td></td>
<td>100,00</td>
<td>2</td>
<td>1,9</td>
<td>5,6</td>
</tr>
<tr>
<td>Gesamt</td>
<td>36</td>
<td>34,6</td>
<td>100,0</td>
<td></td>
</tr>
<tr>
<td>Fehlend</td>
<td>System</td>
<td>68</td>
<td>65,4</td>
<td></td>
</tr>
<tr>
<td>Gesamt</td>
<td></td>
<td>104</td>
<td>100,0</td>
<td></td>
</tr>
</tbody>
</table>

**Ekel visuell richtig erkannt in Prozent**

<table>
<thead>
<tr>
<th></th>
<th>Häufigkeit</th>
<th>Prozent</th>
<th>Gültige Prozente</th>
<th>Kumulierte Prozente</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gültig</td>
<td>0,00</td>
<td>64</td>
<td>61,5</td>
<td>94,1</td>
</tr>
<tr>
<td></td>
<td>100,00</td>
<td>4</td>
<td>3,8</td>
<td>5,9</td>
</tr>
<tr>
<td>Gesamt</td>
<td>68</td>
<td>65,4</td>
<td>100,0</td>
<td></td>
</tr>
<tr>
<td>Fehlend</td>
<td>System</td>
<td>36</td>
<td>34,6</td>
<td></td>
</tr>
<tr>
<td>Gesamt</td>
<td></td>
<td>104</td>
<td>100,0</td>
<td></td>
</tr>
</tbody>
</table>
Ekel ist demnach anhand der Stimme wesentlich deutlicher zu erkennen als anhand einer reinen Audio-Datei.

- **Angst:**

<table>
<thead>
<tr>
<th>Statistiken</th>
<th>Angst auditiv richtig erkannt in Prozent</th>
<th>Angst visuell richtig erkannt in Prozent</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Gültig</td>
<td>33</td>
<td>35</td>
</tr>
<tr>
<td>Fehlend</td>
<td>71</td>
<td>69</td>
</tr>
<tr>
<td>Mittelwert</td>
<td>93,9394</td>
<td>94,2857</td>
</tr>
<tr>
<td>Modus</td>
<td>100,00</td>
<td>100,00</td>
</tr>
<tr>
<td>Standardabweichung</td>
<td>24,23058</td>
<td>23,55041</td>
</tr>
</tbody>
</table>

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:**

<table>
<thead>
<tr>
<th>Statistiken</th>
<th>Wut auditiv richtig erkannt in Prozent</th>
<th>Wut visuell richtig erkannt in Prozent</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Gültig</td>
<td>36</td>
<td>33</td>
</tr>
<tr>
<td>Fehlend</td>
<td>68</td>
<td>71</td>
</tr>
<tr>
<td>Mittelwert</td>
<td>52,7778</td>
<td>78,7879</td>
</tr>
<tr>
<td>Modus</td>
<td>100,00</td>
<td>100,00</td>
</tr>
<tr>
<td>Standardabweichung</td>
<td>50,63094</td>
<td>41,51488</td>
</tr>
</tbody>
</table>

**Neugierde:**

<table>
<thead>
<tr>
<th></th>
<th>Neugierde auditiv richtig erkannt in Prozent</th>
<th>Neugierde visuell richtig erkannt in Prozent</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Gültig</td>
<td>33</td>
<td>35</td>
</tr>
<tr>
<td>Fehlend</td>
<td>71</td>
<td>69</td>
</tr>
<tr>
<td>Mittelwert</td>
<td>39,3939</td>
<td>40,0000</td>
</tr>
<tr>
<td>Modus</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Standardabweichung</td>
<td>49,61977</td>
<td>49,70501</td>
</tr>
</tbody>
</table>


**Überraschung:**

<table>
<thead>
<tr>
<th></th>
<th>Überraschung auditiv richtig erkannt in Prozent</th>
<th>Überraschung visuell richtig erkannt in Prozent</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Gültig</td>
<td>104</td>
<td>71</td>
</tr>
<tr>
<td>Fehlend</td>
<td>0</td>
<td>33</td>
</tr>
<tr>
<td>Mittelwert</td>
<td>25,9615</td>
<td>25,3521</td>
</tr>
<tr>
<td>Modus</td>
<td>.00</td>
<td>.00</td>
</tr>
</tbody>
</table>

**Freude:**

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.

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.

<table>
<thead>
<tr>
<th>Ränge</th>
<th>Mittlerer Rang</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alle Videos - Auditiv richtig erkannt in Prozent</td>
<td>1,75</td>
</tr>
<tr>
<td>Alle Videos - Visuell richtig erkannt in Prozent</td>
<td>2,11</td>
</tr>
<tr>
<td>Alle Videos - Audiovisuell richtig erkannt in Prozent</td>
<td>2,14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistik für Test*</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>104</td>
</tr>
<tr>
<td>Chi-Quadrat</td>
<td>10,784</td>
</tr>
<tr>
<td>df</td>
<td>2</td>
</tr>
</tbody>
</table>
Asymptotische Signifikanz | .005

a. Friedman-Test


### Statistiken

<table>
<thead>
<tr>
<th></th>
<th>Alle Videos - Auditiv richtig erkannt in Prozent</th>
<th>Alle Videos - Visuell richtig erkannt in Prozent</th>
<th>Alle Videos - Audiovisuell richtig erkannt in Prozent</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>104</td>
<td>104</td>
<td>104</td>
</tr>
<tr>
<td>Gültig</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Fehlend</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mittelwert</td>
<td>39,5032</td>
<td>48,5096</td>
<td>49,9199</td>
</tr>
<tr>
<td>Modus</td>
<td>66,67</td>
<td>50,00</td>
<td>66,67</td>
</tr>
<tr>
<td>Standardabweichung</td>
<td>25,01335</td>
<td>16,55939</td>
<td>26,28815</td>
</tr>
</tbody>
</table>

### Alle Videos - Auditiv richtig erkannt in Prozent

<table>
<thead>
<tr>
<th></th>
<th>Häufigkeit</th>
<th>Prozent</th>
<th>Gültige Prozente</th>
<th>Kumulierte Prozente</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gültig</td>
<td>0 % richtig erkannt</td>
<td>12</td>
<td>11,5</td>
<td>11,5</td>
</tr>
<tr>
<td></td>
<td>0,1 bis 25 % richtig erkannt</td>
<td>28</td>
<td>26,9</td>
<td>26,9</td>
</tr>
<tr>
<td></td>
<td>25,1 bis 50 % richtig erkannt</td>
<td>35</td>
<td>33,7</td>
<td>33,7</td>
</tr>
<tr>
<td></td>
<td>50,1 bis 75 % richtig erkannt</td>
<td>25</td>
<td>24,0</td>
<td>24,0</td>
</tr>
<tr>
<td></td>
<td>mehr als 75 % richtig erkannt</td>
<td>4</td>
<td>3,8</td>
<td>3,8</td>
</tr>
<tr>
<td>Gesamt</td>
<td>104</td>
<td>100,0</td>
<td>100,0</td>
<td></td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Ränge</th>
<th>N</th>
<th>Mittlerer Rang</th>
<th>Rangsumme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alle Videos - Audiovisuell richtig erkannt in Prozent - Alle Videos - Auditiv richtig erkannt in Prozent</td>
<td>34⁹</td>
<td>44,93</td>
<td>1527,50</td>
</tr>
<tr>
<td>Positive Ränge</td>
<td>58⁹</td>
<td>47,42</td>
<td>2750,50</td>
</tr>
<tr>
<td>Bindungen</td>
<td>12⁹</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gesamt</td>
<td>104</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 Test⁹

| Z | -2,391⁹ |
| 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.
<table>
<thead>
<tr>
<th>Bindungen</th>
<th>Gesamt</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>104</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Statistik für Test&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alle Videos - Audiovisuell richtig erkannt in Prozent - Alle Videos - Visuell richtig erkannt in Prozent</td>
</tr>
<tr>
<td>Z</td>
</tr>
<tr>
<td>Asymptotische Signifikanz (2-seitig)</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Statistiken</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traurigkeit audiovisuell richtig erkannt in Prozent</td>
</tr>
<tr>
<td>N Gültig</td>
</tr>
<tr>
<td>Fehlend</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>N Gültig</td>
</tr>
<tr>
<td>Fehlend</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th></th>
<th>Ekel audiovisuell richtig erkannt in Prozent</th>
<th>Traurigkeit audiovisuell richtig erkannt in Prozent</th>
<th>Vertrauen audiovisuell richtig erkannt in Prozent</th>
<th>Angst audiovisuell richtig erkannt in Prozent</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Gültig</td>
<td>68</td>
<td>68</td>
<td>68</td>
<td>36</td>
</tr>
<tr>
<td>Fehlend</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>68</td>
</tr>
<tr>
<td>Mittelwert</td>
<td>33,8235</td>
<td>52,9412</td>
<td>39,7059</td>
<td>91,6667</td>
</tr>
<tr>
<td>Modus</td>
<td>.00</td>
<td>100,00</td>
<td>.00</td>
<td>100,00</td>
</tr>
<tr>
<td>Standardabweichung</td>
<td>47,66266</td>
<td>50,28453</td>
<td>49,29263</td>
<td>28,03060</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Wut audiovisuell richtig erkannt in Prozent</th>
<th>Neugierde audiovisuell richtig erkannt in Prozent</th>
<th>Freude audiovisuell richtig erkannt in Prozent</th>
<th>Überraschung audiovisuell richtig erkannt in Prozent</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Gültig</td>
<td>35</td>
<td>36</td>
<td>33</td>
<td>69</td>
</tr>
<tr>
<td>Fehlend</td>
<td>69</td>
<td>68</td>
<td>71</td>
<td>35</td>
</tr>
<tr>
<td>Mittelwert</td>
<td>77,1429</td>
<td>25,0000</td>
<td>21,2121</td>
<td>34,0580</td>
</tr>
<tr>
<td>Modus</td>
<td>100,00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Standardabweichung</td>
<td>42,60430</td>
<td>43,91550</td>
<td>41,51488</td>
<td>42,43746</td>
</tr>
</tbody>
</table>
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
audiovisuellen Ausschnitte. Beide Formen sind bei negativen Emotionen jedoch der stimmlichen Äußerung als Kommunikationsweg überlegen.


 Aus diesen Ergebnissen lässt sich folgende neue Hypothese generieren: 

**Hypothese 4.4:** Emotionen werden anhand audiovisueller Ausschnitte in ähnlicherem 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**

<table>
<thead>
<tr>
<th></th>
<th>Alle Videos - Auditiv richtig erkannt in Prozent</th>
<th>FPI-R Extraversion Statine</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spearman-Rho</strong></td>
<td><strong>FPI-R Extraversion Statine</strong></td>
<td></td>
</tr>
<tr>
<td>Alle Videos - Auditiv richtig erkannt in Prozent</td>
<td>Korrelationskoeffizient 1,000</td>
<td>-0,271**</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-seitig)</td>
<td>.005</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>104</td>
</tr>
<tr>
<td>FPI-R Extraversion Statine</td>
<td>Korrelationskoeffizient -0,271**</td>
<td>1,000</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-seitig)</td>
<td>.005</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>104</td>
</tr>
</tbody>
</table>

**. 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**

<table>
<thead>
<tr>
<th>Fällen</th>
<th>Fälle</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gültig</td>
<td>Fehlend</td>
<td>Gesamt</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>Prozent</td>
<td>N</td>
<td>Prozent</td>
</tr>
<tr>
<td>FPI-R Extraversion Statine *</td>
<td>Alle Videos - Auditiv richtig erkannt - Gruppen Prozent</td>
<td>104</td>
<td>100,0%</td>
<td>0</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Anzahl</th>
<th>Alle Videos - Auditiv richtig erkannt - Gruppen Prozent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 % richtig erkannt</td>
</tr>
<tr>
<td>FPI-R Extraversion Statine 1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>
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**

<table>
<thead>
<tr>
<th></th>
<th>Gültig</th>
<th>Fehlend</th>
<th>Mittelwert</th>
<th>Standardabweichung</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPI-R Extraversion Statine</td>
<td></td>
<td></td>
<td>4,67</td>
<td>1,692</td>
</tr>
</tbody>
</table>

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**

<table>
<thead>
<tr>
<th></th>
<th>FPI-R Extraversion Statine</th>
<th>Alle Videos - Visuell richtig erkannt in Prozent</th>
<th>Alle Videos - Audiovisuell richtig erkannt in Prozent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman-Rho FPI-R Extraversion Statine</td>
<td>Korrelationskoeffizient</td>
<td>-0,13</td>
<td>0,164</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-seitig)</td>
<td>0,90</td>
<td>0,095</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>104</td>
<td>104</td>
</tr>
<tr>
<td>Alle Videos - Visuell richtig erkannt in Prozent</td>
<td>Korrelationskoeffizient</td>
<td>-0,13</td>
<td>1,000</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-seitig)</td>
<td>0,900</td>
<td>0,072</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>104</td>
<td>104</td>
</tr>
<tr>
<td>Alle Videos - Audiovisuell richtig erkannt in Prozent</td>
<td>Korrelationskoeffizient</td>
<td>0,164</td>
<td>0,177</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-seitig)</td>
<td>0,095</td>
<td>0,072</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>104</td>
<td>104</td>
</tr>
</tbody>
</table>

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

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.
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.

<table>
<thead>
<tr>
<th>Korrelationen</th>
<th>Alle Videos - Auditiv richtig erkannt in Prozent</th>
<th>FPI-R Gehemmtheit Statine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman-Rho</td>
<td>Alle Videos - Auditiv richtig erkannt in Prozent</td>
<td>Korrelationskoeffizient</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sig. (2-seitig)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>FPI-R Gehemmtheit Statine</td>
<td>Korrelationskoeffizient</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sig. (2-seitig)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Korrelationen</th>
<th>Alle Videos - Auditiv richtig erkannt in Prozent</th>
<th>Empathie in realen Situationen - Punktedurchschnitt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman-Rho</td>
<td>Alle Videos - Auditiv richtig erkannt in Prozent</td>
<td>Korrelationskoeffizient</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sig. (2-seitig)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Empathie in realen Situationen - Punktedurchschnitt</td>
<td>Korrelationskoeffizient</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sig. (2-seitig)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th><strong>Korrelationen</strong></th>
<th>Alle Videos - Auditiv richtig erkannt in Prozent</th>
<th>Phantasieempathie - Punktedurchschnitt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman-Rho</td>
<td>Korrelationskoeffizient</td>
<td>1,000</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-seitig)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>104</td>
</tr>
<tr>
<td>Phantasieempathie</td>
<td>Korrelationskoeffizient</td>
<td>.012</td>
</tr>
<tr>
<td>Punktedurchschnitt</td>
<td>Sig. (2-seitig)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>104</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th><strong>Korrelationen</strong></th>
<th>Alle Videos - Auditiv richtig erkannt in Prozent</th>
<th>Geschlecht</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman-Rho</td>
<td>Korrelationskoeffizient</td>
<td>1,000</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-seitig)</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>104</td>
</tr>
<tr>
<td>Geschlecht</td>
<td>Korrelationskoeffizient</td>
<td>-.004</td>
</tr>
</tbody>
</table>
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.

Tabelle 6.2: Korrelationen

<table>
<thead>
<tr>
<th></th>
<th>Alle Videos - Auditiv richtig erkannt in Prozent</th>
<th>Alter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman-Rho</td>
<td>Korrelationskoeffizient</td>
<td>1,000</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-seitig)</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>104</td>
</tr>
<tr>
<td>Alter</td>
<td>Korrelationskoeffizient</td>
<td>-0,24</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-seitig)</td>
<td>0,810</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>103</td>
</tr>
</tbody>
</table>

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.

Tabelle 6.3: Korrelationen

<table>
<thead>
<tr>
<th></th>
<th>Alle Videos - Auditiv richtig erkannt in Prozent</th>
<th>Haben Sie Geschwister?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman-Rho</td>
<td>Korrelationskoeffizient</td>
<td>1,000</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-seitig)</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>104</td>
</tr>
<tr>
<td>Haben Sie Geschwister?</td>
<td>Korrelationskoeffizient</td>
<td>0,040</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-seitig)</td>
<td>0,686</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>103</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Korrelationen</th>
<th>Alle Videos - Auditiv richtig erkannt in Prozent</th>
<th>Haben Sie Kinder?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman-Rho</td>
<td>Korrelationskoeffizient 1,000</td>
<td>-.065</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-seitig) .</td>
<td>.512</td>
</tr>
<tr>
<td></td>
<td>N 104</td>
<td>104</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Korrelationen</th>
<th>Alle Videos - Auditiv richtig erkannt in Prozent</th>
<th>Haben Sie Kinder? - Gruppen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman-Rho</td>
<td>Korrelationskoeffizient 1,000</td>
<td>-.131</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-seitig) .</td>
<td>.184</td>
</tr>
<tr>
<td></td>
<td>N 104</td>
<td>104</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Haben Sie Kinder? - Gruppen * Alle Videos - Auditiv richtig erkannt - Gruppen Prozent Kreuztabelle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anzahl</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alle Videos - Auditiv richtig erkannt - Gruppen Prozent</th>
<th>Gesamt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haben Sie Kinder? - Gruppen</td>
<td>Keine Kinder</td>
</tr>
<tr>
<td>----------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>0 % richtig erkannt</td>
<td>6</td>
</tr>
<tr>
<td>0,1 bis 25 % richtig erkannt</td>
<td>18</td>
</tr>
<tr>
<td>25,1 bis 50 % richtig erkannt</td>
<td>19</td>
</tr>
<tr>
<td>50,1 bis 75 % richtig erkannt</td>
<td>15</td>
</tr>
<tr>
<td>mehr als 75 % richtig erkannt</td>
<td>3</td>
</tr>
<tr>
<td><strong>Gesamt</strong></td>
<td><strong>12</strong></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Korrelationen</th>
<th>Alle Videos - Auditiv richtig erkannt in Prozent</th>
<th>Familienstand - Gruppen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman-Rho</td>
<td>Korrelationskoeffizient</td>
<td>Sig. (2-seitig)</td>
</tr>
<tr>
<td>Alle Videos - Auditiv richtig erkannt in Prozent</td>
<td>1,000</td>
<td>.153</td>
</tr>
<tr>
<td>Sig. (2-seitig)</td>
<td>.</td>
<td>.121</td>
</tr>
<tr>
<td>N</td>
<td>104</td>
<td>104</td>
</tr>
</tbody>
</table>

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**
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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N Gültig</td>
<td>104</td>
<td>104</td>
<td>99</td>
</tr>
<tr>
<td>Fehlend</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Mittelwert</td>
<td>2,87</td>
<td>1,07</td>
<td>1,60</td>
</tr>
<tr>
<td>Standardabweichung</td>
<td>1,637</td>
<td>0,349</td>
<td>1,087</td>
</tr>
</tbody>
</table>

Korrelationen

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman-Rho Haben Sie „CSI: Miami“ schon einmal gesehen?</td>
<td>Korrelationskoeffizient</td>
<td>1,000</td>
<td>-0,168</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-seitig)</td>
<td>.160</td>
<td>.958</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>104</td>
<td>104</td>
</tr>
<tr>
<td>CSI: Miami - Audio richtig erkannt</td>
<td>Korrelationskoeffizient</td>
<td>-0,168</td>
<td>1,000</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-seitig)</td>
<td>.160</td>
<td>.958</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>71</td>
<td>71</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>CSI: Miami - Audio richtig erkannt</th>
<th>Serienkenntnis-Erhebung - Sehen Sie „CSI: Miami“ regelmäßig im Fernsehen?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman-Rho</td>
<td>Korrelationskoeffizient</td>
<td>Korrelationskoeffizient</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-seitig)</td>
<td>Sig. (2-seitig)</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>1,000</td>
<td>-0,081</td>
</tr>
<tr>
<td></td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>71</td>
<td>68</td>
</tr>
</tbody>
</table>

### Korrelationen

<table>
<thead>
<tr>
<th></th>
<th>Serienkenntnis-Erhebung - Sahen Sie „Damages“ regelmäßig im Fernsehen?</th>
<th>Damages - Audio richtig erkannt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman-Rho</td>
<td>Korrelationskoeffizient</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sig. (2-seitig)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

### Statistiken

<table>
<thead>
<tr>
<th></th>
<th>Serienkenntnis-Erhebung - Sahen Sie „Damages“ regelmäßig im Fernsehen?</th>
<th>Damages - Audio richtig erkannt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 Hö rer. Sie sehen fern.«


---

280 Vgl. Postman (2006), S. 92


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

\textsuperscript{281} Vgl. Scherer, in: Scherer (1982), S. 303
\textsuperscript{282} Vgl. Merten (2003), S. 153

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 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\textsuperscript{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.


\textsuperscript{284} Vgl. Williams / Stevens, in Scherer (1982), S. 313 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, 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.“

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. 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

\[287\] Vgl. Ebenda, S. 195 f.
4. Literatur:


Zeitschriften:


Kreifelts, Benjamin et al.: *Audiovisual integration of emotional signals in voice and face: An event-related fMRI study*. In: *NeuroImage* (2007), Heft 37. URL: 

Wildgruber, Dirk et al.: *Identification of emotional intonation evaluated by fMRI*. In: *NeuroImage* (2005), Heft 24. URL: 

**Online-Dokumente:**

BBC Online (2006): *CSI show 'most popular in the world'* . URL:  
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.


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)**


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

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Zeit</th>
<th>Kameraperspektive</th>
<th>Einstellungsgröße</th>
<th>Inhalt</th>
<th>Dialog</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>15:50</td>
<td>Leichte Draufsicht</td>
<td>Halbtotale</td>
<td>Ellen sitzt mit ihrem Brautkleid auf dem Bett und weint. Ihre Mutter geht auf sie zu und fragt: „Elli, was ist denn los?“</td>
<td>„Oh… Willst du mein Kleid nicht anziehen?“</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>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:</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>16:02</td>
<td>Normalsicht</td>
<td>Nahe</td>
<td>Ellen lächelt gequält, weint aber weiter, kuschelt sich an ihre Mutter und sagt: „Ich bin so froh, dass du da bist!“</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Die Mutter nimmt sie in den Arm und beruhigt sie:</td>
<td>„Oooh, das bin ich auch, Liebes. Das bin ich doch auch.“</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Zeit</th>
<th>Sicht</th>
<th>Szene</th>
<th>Dialog</th>
<th>Zusätzliche Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>20:30</td>
<td>Normalsicht</td>
<td>Großaufnahme</td>
<td>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.)</td>
<td>„Und dann fängt sie auch noch mit Katie an!“</td>
</tr>
<tr>
<td>20:31</td>
<td>Leichte Untersicht</td>
<td>Halbnah</td>
<td>Ellen nimmt gerade ihre Ohrringe ab und entgegnet:</td>
<td>„Weil Patty sich Sorgen um sie macht!“</td>
</tr>
<tr>
<td>20:33</td>
<td>Normalsicht</td>
<td>Großaufnahme</td>
<td>David dreht sich weg:</td>
<td>„Schwachsinn!“</td>
</tr>
<tr>
<td>20:34</td>
<td>Leichte Untersicht</td>
<td>Halbnah</td>
<td>Ellen wendet wütend ein:</td>
<td>„Du hast dich schon aufgeführt, als wolltest du nicht dort sein, lange bevor Patty…“</td>
</tr>
<tr>
<td>20:38</td>
<td>Normalsicht</td>
<td>Nahe</td>
<td>David hängt gerade sein Sakko in den Schrank… Katie erwähnt hat!“</td>
<td>„Ja, das hast du auch verdammt deutlich…“</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>David bekräftigt:</td>
<td>„So war’s ja auch.“</td>
</tr>
<tr>
<td>20:39</td>
<td>Leichte Untersicht</td>
<td>Halbnah</td>
<td>Ellen sagt:</td>
<td>„Hör zu, es tut mir leid wenn ich mich nicht für…“</td>
</tr>
<tr>
<td>20:41</td>
<td>Normalsicht</td>
<td>Nahe</td>
<td>Blick auf David… gezeigt!“</td>
<td>„… international“</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Die Kamera fährt mit David nach links, während er spricht:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>David geht an Ellen vorbei nach rechts, sie sind</td>
<td></td>
</tr>
<tr>
<td>Zeit</td>
<td>Draufsicht</td>
<td>Sicht</td>
<td>Blick</td>
<td>Äußerungen</td>
</tr>
<tr>
<td>-------</td>
<td>------------</td>
<td>-------------</td>
<td>------------</td>
<td>----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>8:46</td>
<td>Leichte</td>
<td>Halbtotale</td>
<td>Blick vom Badezimmer aus durch die geöffnete Tür zur Kommode, vor der Ellen und David stehen und streiten</td>
<td>„Phil wollte sich nur unterhalten! Und als sich Patty nach dem Krankenhaus erkundigte, …“</td>
</tr>
<tr>
<td>8:51</td>
<td>Normalsicht</td>
<td>Nahe</td>
<td>Ellen entgegnet: (Phil ist Pattys Ehemann, David arbeitet in einem Krankenhaus)</td>
<td>„Nein, du hast gegrunzt wie ein Höhlenmensch!“</td>
</tr>
<tr>
<td>8:51</td>
<td>Normalsicht</td>
<td>Nahe</td>
<td>David unterbricht sie:</td>
<td>„Ich hab’ ihr geantwortet, oder?“</td>
</tr>
<tr>
<td>8:51</td>
<td>Normalsicht</td>
<td>Nahe</td>
<td>Ellen widerspricht energisch:</td>
<td>„Ich musste mich für dich entschuldigen!“</td>
</tr>
<tr>
<td>8:51</td>
<td>Normalsicht</td>
<td>Nahe</td>
<td>David entgegnet widerspenstig:</td>
<td>„Du hast was? Oh!“</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Zeit</th>
<th>Draufsicht</th>
<th>Sicht</th>
<th>Blick</th>
<th>Äußerungen</th>
</tr>
</thead>
</table>
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

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Zeit</th>
<th>Kameraperspektive</th>
<th>Einstellungsgröße</th>
<th>Inhalt</th>
<th>Dialog</th>
</tr>
</thead>
</table>
Als Textinsert in der deutschsprachigen Fassung ist zu lesen: „Milliardär flippt aus – Arthur Frobisher der Körperverletzung beschuldigt“

<table>
<thead>
<tr>
<th>Zeile</th>
<th>Normalsicht</th>
<th>Medium Shot</th>
<th>Die Kamera schwenkt langsam nach links oben, währenddessen kommt Patty Hewes (Ellens Chefin) von hinten ins Bild, man hört Ellens Stimme, die vorliest:</th>
<th>„Milliardär flippt aus – Arthur Frobisher der Körperverletzung beschuldigt.“</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>32:50</td>
<td>Normalsicht</td>
<td>Großaufnahme</td>
<td>Ellen sieht Patty an und fragt: „Werden sie ihn festnehmen?“</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ellen erwidert: „Nein, David arbeitet, also bleibe ich heute länger.“</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Patty sieht sie skeptisch an und fragt: „Ich hab’ über David und Sie nachgedacht. Wollen Sie einen Drink?“</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>33:09</td>
<td>Normalsicht</td>
<td>Großaufnahme</td>
<td>Ellen sieht Patty an und lächelt unsicher: „Gern.“</td>
</tr>
<tr>
<td>5</td>
<td>33:11</td>
<td>Normalsicht</td>
<td>Medium Shot</td>
<td>Patty lächelt Ellen an und dreht sich um in Richtung Tür</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>6</td>
<td>33:12</td>
<td>Normalsicht</td>
<td>Großaufnahme</td>
<td>Ellen steht auf und geht hinter Patty her</td>
</tr>
<tr>
<td>7</td>
<td>33:14</td>
<td>Leichte Draufsicht</td>
<td>Halbtotale</td>
<td>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.“</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Leichte Draufsicht</td>
<td>Großaufnahme</td>
<td>Die Kamera bleibt auf Ellens Gesicht stehen als sie überrascht fragt: „Warum sollten wir das nicht?“</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Amerikanische Einstellung</td>
<td>Die Kamera schwenkt weiter auf Ellen und Patty, die eine Schranktür öffnet und sagt: „Es ist verdammt hart…“</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>33:21</td>
<td>Leichte Untersicht</td>
<td>Nahe</td>
<td>Ellen steht neben einer niedrigen Lampe, Patty spricht weiter: „… Ellen, Mann und Karriere unter einen Hut zu bringen.“</td>
</tr>
<tr>
<td>9</td>
<td>33:24</td>
<td>Leichte Draufsicht</td>
<td>Halbnah</td>
<td>Patty hält Ellen grinsend einen Drink entgegen: „Cheers!“</td>
</tr>
<tr>
<td>10</td>
<td>33:26</td>
<td>Leichte Untersicht</td>
<td>Nahe</td>
<td>Ellen nimmt Patty lächelnd das Glas aus der Hand, sieht hinein und sagt ebenfalls: „Cheers!“</td>
</tr>
<tr>
<td>Zeit</td>
<td>Leichte Unterk.</td>
<td>Nähe</td>
<td>Szene/Credit</td>
<td>Deutsch</td>
</tr>
<tr>
<td>-------</td>
<td>----------------</td>
<td>-------</td>
<td>--------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>12</td>
<td>33:43</td>
<td>Nahe</td>
<td>Patty sitzt neben ihr und sieht sie an. Patty spricht weiter:</td>
<td>„Er hatte zwei Jobs, ganz undankbare, niedere…</td>
</tr>
<tr>
<td>13</td>
<td>33:47</td>
<td>Halbnah</td>
<td>Schnitt auf Patty</td>
<td>… Tätigkeiten. Und… er war meine erste große Liebe.“</td>
</tr>
<tr>
<td>14</td>
<td>33:54</td>
<td>Nahe</td>
<td>Ellen sitzt neben Patty, die einen Schluck nimmt, und fragt sie:</td>
<td>„Ich habe ihn nach dem Examen abserviert.“</td>
</tr>
<tr>
<td>15</td>
<td>33:56</td>
<td>Halbnah</td>
<td>Schnitt auf Patty, die eiskalt erwidert:</td>
<td>„Er hatte keinerlei Ehrgeiz. Das hat mich angewidert.“</td>
</tr>
<tr>
<td>16</td>
<td>34:00</td>
<td>Nahe</td>
<td>Patty sitzt nickend neben Ellen, die etwas fassungslos fragt:</td>
<td>„Im Ernst?“</td>
</tr>
<tr>
<td>17</td>
<td>34:02</td>
<td>Halbnah</td>
<td>Patty sieht Ellen mit einem Lächeln an</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>34:04</td>
<td>Nahe</td>
<td>Ellen sieht Patty ungläubig an</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>34:06</td>
<td>Halbnah</td>
<td>Patty antwortet:</td>
<td>„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.“</td>
</tr>
<tr>
<td>Zeit</td>
<td>Sicht</td>
<td>Nähe</td>
<td>Text</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
<td>------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>20:34:15</td>
<td>Leichte Untersicht</td>
<td>Nahe</td>
<td>Ellen nickt und fragt: (Phil ist Pattys Ehemann) „Hat Phil deswegen beim Essen so viel geredet?“</td>
<td></td>
</tr>
<tr>
<td>21:34:18</td>
<td>Leichte Draufsicht</td>
<td>Halbnah</td>
<td>Patty schweigt erst, dann lacht sie</td>
<td></td>
</tr>
<tr>
<td>22:34:22</td>
<td>Leichte Untersicht</td>
<td>Nahe</td>
<td>Ellen sieht sie etwas verständnislos an, mit einem halben Lächeln, Patty lacht weiter</td>
<td></td>
</tr>
<tr>
<td>23:34:24</td>
<td>Leichte Draufsicht</td>
<td>Halbnah</td>
<td>Patty meint lachend: „Ja, er hält gerne …</td>
<td></td>
</tr>
<tr>
<td>24:34:26</td>
<td>Leichte Untersicht</td>
<td>Nahe</td>
<td>Ellen sieht nachdenklich nach unten … Hof!“</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Patty blickt Ellen an und fragt: „Was?“</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ellen erklärt: „Ich habe David nicht erzählt, dass ich mich mit Gregory … getroffen hab‘.“</td>
<td></td>
</tr>
<tr>
<td>25:34:35</td>
<td>Leichte Draufsicht</td>
<td>Halbnah</td>
<td>Patty sieht sie an: „Wirklich?“</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ellen erklärt unsicher: „Ich schätze, ich wollte ihn nicht beunruhigen.“</td>
<td></td>
</tr>
<tr>
<td>26:34:38</td>
<td>Leichte Untersicht</td>
<td>Nahe</td>
<td>Patty erwidert: „Ellen… Die meisten Männer können nicht mit ehrgeizigen Frauen umgehen.“</td>
<td></td>
</tr>
</tbody>
</table>
| 27:34:41 | Leichte Draufsicht | Halbnahe | 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 … dauern, aber bemühen Sie sich,
Damages – „David!“ (Angst)
Beginn: 01:13 – Ende: 01:28

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

<table>
<thead>
<tr>
<th>Zeit</th>
<th>Aufnahmeart</th>
<th>Perspektive</th>
<th>Beschreibung</th>
</tr>
</thead>
<tbody>
<tr>
<td>01:18</td>
<td>Untersicht</td>
<td>Detailaufnahme</td>
<td>Blick auf Davids blutige Hand, die ebenfalls aus der Badewanne hängt. Im Hintergrund ist verschwommen sein blutiges Gesicht zu erkennen</td>
</tr>
<tr>
<td>01:19</td>
<td>Leichte Untersicht</td>
<td>Nahe</td>
<td>Schnitt auf Davids Gesicht und sein weißes, blutbeflecktes T-Shirt</td>
</tr>
<tr>
<td>01:19</td>
<td>Leichte Untersicht</td>
<td>Detailaufnahme</td>
<td>Ellen lässt die blutige Statuette aus ihrer Hand fallen</td>
</tr>
<tr>
<td>01:20</td>
<td>Normalsicht</td>
<td>Nahe</td>
<td>Blick von hinten auf Ellen, die vor der Wanne steht – gerade fällt die kleine Statue zu Boden</td>
</tr>
<tr>
<td>01:21</td>
<td>Draufsicht</td>
<td>Halbtotale</td>
<td>Blick von der Zimmerdecke auf die Badewanne, Ellen stürzt zu ihrem toten Verlobten nieder</td>
</tr>
<tr>
<td>01:23</td>
<td>Normalsicht</td>
<td>Nahe</td>
<td>Ellen nimmt Davids Kopf in ihre Arme und hält ihn fest</td>
</tr>
</tbody>
</table>
CSI: Miami – „Ich ruf’ da jetzt nicht an!“ (Überraschung, Abscheu/Ekel)
Beginn: 06:18 – Ende: 07:07

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Zeit</th>
<th>Kameraperspektive</th>
<th>Einstellungsgröße</th>
<th>Inhalt</th>
<th>Dialog</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>06:18</td>
<td>Draufsicht</td>
<td>Detailaufnahme</td>
<td>Kreisrunder Blick auf gelbes Material</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>06:21</td>
<td>Normalsicht</td>
<td>Nah</td>
<td>Tim Speedle sieht in Mikroskop</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>06:22</td>
<td>Leichte Draufsicht</td>
<td>Establishing Shot</td>
<td>Eric Delko betritt das Labor, geht auf Speedle (der rechts im Bild sitzt) zu und spricht ihn an</td>
<td>„Es ist ein Lamborghini.“</td>
</tr>
<tr>
<td>4</td>
<td>06:24</td>
<td>Leichte Draufsicht</td>
<td>Halbtotale</td>
<td>Kamera fährt nach unten und auf Delko und Speedle zu</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>06:27</td>
<td>Leichte Untersicht</td>
<td>Medium Shot</td>
<td>Delko spricht weiter</td>
<td>„Der Farbcode ist 0056.“</td>
</tr>
<tr>
<td>6</td>
<td>06:30</td>
<td>Normalsicht</td>
<td>Nahe</td>
<td>Speedle sieht sein Gegenüber an, von dem nur ein Teil des Arms im Bild ist, und sagt:</td>
<td>„Hellgelb metallic“</td>
</tr>
<tr>
<td>6</td>
<td>06:32</td>
<td>Leichte Untersicht</td>
<td>Nahe</td>
<td>Blick auf Delko, der sich überrascht zeigt:</td>
<td>„Genau!“</td>
</tr>
<tr>
<td>7</td>
<td>06:33</td>
<td>Normalsicht</td>
<td>Nahe</td>
<td>Schnitt auf Speedle:</td>
<td>„Was denkst du, wie viele Leute hier so ein Auto fahren?“</td>
</tr>
<tr>
<td>8</td>
<td>06:36</td>
<td>Leichte Untersicht</td>
<td>Nahe</td>
<td>Wieder Blick auf Delko:</td>
<td>„Vielleicht hat er’s gemietet. Der kostet ,ne Viertelmillion Dollar, da muss ,ne alte Frau lang für stricken.“</td>
</tr>
<tr>
<td>9</td>
<td>06:40</td>
<td>Normalsicht</td>
<td>Nahe</td>
<td>Speedle sieht auf etwas rechts von ihm (nicht</td>
<td>„Naja, das schränkt die Suche</td>
</tr>
<tr>
<td>Zeit</td>
<td>Normal</td>
<td>Aufnahme</td>
<td>Sprecher</td>
<td>Text</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>------------</td>
<td>-----------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>10:06:42</td>
<td>10:06:42</td>
<td>Speedle hebt den Telefonhörer auf:</td>
<td>„Hast du noch Kontakt zu der Maus, die bei der Zulassung arbeitet?“</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:06:46</td>
<td>10:06:46</td>
<td>Delko sieht überrascht aus:</td>
<td>„Gina? Ja, wieso?“</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:06:47</td>
<td>10:06:47</td>
<td>Speedle reicht Delko den Telefonhörer:</td>
<td>„Hier!“</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:06:49</td>
<td>10:06:49</td>
<td>Delko verzieht das Gesicht:</td>
<td>„Ich ruf’ da jetzt nicht an! Das ist problematisch!“</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:06:55</td>
<td>10:06:55</td>
<td>Delko sieht Speedle an, dessen Hand mit Telefonhörer unten im Bild zu sehen ist, und sagt:</td>
<td>„Darauf hab’ ich keine Lust!“</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:06:57</td>
<td>10:06:57</td>
<td>Speedle gibt nicht nach:</td>
<td>„Eric, ab und zu musst du auch mal mit…“</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:06:58</td>
<td>10:06:58</td>
<td>Blick auf Delko, Speedle spricht weiter:</td>
<td>„… denen ausgehen. Beziehungen muss man pflegen!“</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:07:00</td>
<td>10:07:00</td>
<td>Schnitt auf Speedle:</td>
<td>„Du rufst sie jetzt an! Ich wähl’ auch für dich.“</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:07:04</td>
<td>10:07:04</td>
<td>Delko verzieht wieder das Gesicht, presst die</td>
<td>„Toll.“</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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)**


<table>
<thead>
<tr>
<th>Nr.</th>
<th>Zeit</th>
<th>Kameraperspektive</th>
<th>Einstellungsgröße</th>
<th>Inhalt</th>
<th>Dialog</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25:15</td>
<td>Leichte Draufsicht</td>
<td>Supertotale</td>
<td>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.</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>25:27</td>
<td>Normalsicht</td>
<td>Detailaufnahme</td>
<td>Es sind Beine zu sehen, die durch Müllberge waten, außerdem greifen Hände in Plastikhandschuhen nach dem Müll und lassen ihn wieder fallen</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>25:29</td>
<td>Normalsicht</td>
<td>Amerikanische Einstellung</td>
<td>Eric Delko hebt im Schutzanzug Müll auf und wirft ihn wieder weg, die Kamera schwenkt nach links unten, wo Tim Speedle das Gleiche tut</td>
<td></td>
</tr>
<tr>
<td>Zeit</td>
<td>Drauf-</td>
<td>Vorschau</td>
<td>Bemerkung</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>----------</td>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 25:32</td>
<td>Leichte Draufsicht</td>
<td>Totale</td>
<td>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</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 25:36</td>
<td>Normalsicht</td>
<td>Amerikanische Einstellung</td>
<td>Speedle sucht den Müll ab, Schwenk nach rechts auf Delko, der sich nach unten beugt und ausruft: „Hey!“</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 25:39</td>
<td>Leichte Untersicht</td>
<td>Halbnah</td>
<td>Speedle sieht nach unten in Richtung Delko, der aus dem Hintergrund ruft: „Sieh dir das hier mal an!“</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 25:41</td>
<td>Normalsicht</td>
<td>Detailaufnahme</td>
<td>Delko hebt etwas auf – eine tote Ratte kommt groß ins Bild</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 25:43</td>
<td>Leichte Untersicht</td>
<td>Halbnah</td>
<td>Speedle blickt auf</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 25:44</td>
<td>Leichte Untersicht</td>
<td>Halbnah</td>
<td>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.“</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 25:48</td>
<td>Leichte Untersicht</td>
<td>Halbnah</td>
<td>Speedle entgegnet mit genervtem Gesichtsausdruck: „Haha, witzig.“</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 25:50</td>
<td>Leichte Untersicht</td>
<td>Halbnah</td>
<td>Delko steht mit ausgestreckter Hand da, wirft die Ratte aber doch weg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 25:51</td>
<td>Leichte Untersicht</td>
<td>Halbnah</td>
<td>Speedle sucht weiter, im Vordergrund fliegt die Ratte durchs Bild. Speedle seufzt: „Puh.“</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zeit</td>
<td>Normalsicht/Leichte Untersicht</td>
<td>Bildschärfe</td>
<td>Beschreibung</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------------------------------</td>
<td>-------------</td>
<td>--------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25:54</td>
<td>Normalsicht</td>
<td>Totale</td>
<td>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.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25:59</td>
<td>Leichte Untersicht</td>
<td>Amerikanische Einstellung</td>
<td>Die Kamera fährt mit als sich Speedle nach unten beugt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26:00</td>
<td>Normalsicht</td>
<td>Detaillaufnahme</td>
<td>Speedle ist links im Bild zu sehen, seine Hände wühlen sich durch den Müll</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26:02</td>
<td>Normalsicht</td>
<td>Halbnah</td>
<td>Kamera schwenkt mit als Speedle etwas aus dem Müll zieht</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26:05</td>
<td>Normalsicht</td>
<td>Nahe</td>
<td>Speedle zieht weiter, er hat plötzlich ein schmerzverzerrtes Gesicht</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26:06</td>
<td>Leichte Untersicht</td>
<td>Amerikanische Einstellung</td>
<td>Speedle greift sich mit der rechten Hand auf den Rücken, gleichzeitig hält er etwas in der linken Hand, er jammert: „Au!“</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26:07</td>
<td>Leichte Untersicht</td>
<td>Halbnah</td>
<td>Im Vordergrund ist verschwommen Speedle mit dem Gegenstand in der Hand zu sehen, im Hintergrund rechts steht Delko. Die Kamera</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Delko streckt die Hände nach dem Gegenstand aus und fragt: „Ist es das?“

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Zeit</th>
<th>Kameraperspektive</th>
<th>Einstellungsgröße</th>
<th>Inhalt</th>
<th>Dialog</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18:22</td>
<td>Leichte Unsicht</td>
<td>Establishing Shot</td>
<td>Kamera schwenkt nach links Horatio Caine und Yelina Salas gehen aufeinander zu, sie bleiben im hinteren Bildteil stehen</td>
<td>„Ich vermute, das State Department hält die Hand über Clavo.“</td>
</tr>
<tr>
<td>2</td>
<td>18:26</td>
<td></td>
<td></td>
<td>Caine lacht, Salas sieht ihn an und sagt:</td>
<td>„Ja, das ist unser Airbag.“</td>
</tr>
<tr>
<td>Zeitaufnahme</td>
<td>Sicht</td>
<td>Einstellung</td>
<td>Handlung</td>
<td>Sprechende Person(en)</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------</td>
<td>-------------</td>
<td>----------</td>
<td>----------------------</td>
<td></td>
</tr>
<tr>
<td>3 18:31</td>
<td>Leichte Untersicht</td>
<td>Amerikanische Einstellung</td>
<td>Salas fragt:</td>
<td>„Was jetzt? Hören wir auf?“</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Salas:</td>
<td>„Also weiter.“</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Caine:</td>
<td>„Mhm… Ist das nicht Statler da drüben?“</td>
<td></td>
</tr>
<tr>
<td>4 18:44</td>
<td>Leichte Untersicht</td>
<td>Medium Shot</td>
<td>Blick auf einen Mann im Nebenraum, der an einem Bartresen sitzt; Salas:</td>
<td>„Ja, er meinte…“</td>
<td></td>
</tr>
<tr>
<td>5 18:47</td>
<td>Normalsicht</td>
<td>Over-the-Shoulder Close-Up</td>
<td>Salas blickt Caine an und spricht weiter:</td>
<td>„… du hättest gesagt, er solle mit…“</td>
<td></td>
</tr>
<tr>
<td>6 18:48</td>
<td>Normalsicht</td>
<td>Over-the-Shoulder Close-Up</td>
<td>Caine sieht Salas an, sie redet weiter:</td>
<td>„… mir reden.“</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Caine antwortet:</td>
<td>„Nein, das ist nicht wahr. Ich habe ihm gesagt, ich kann nicht“</td>
<td></td>
</tr>
<tr>
<td>Zeit</td>
<td>Normalsicht</td>
<td>Übernahme</td>
<td>Scene</td>
<td>Sprache</td>
<td>Inhaltsangabe</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>-----------</td>
<td>-------</td>
<td>---------</td>
<td>---------------</td>
</tr>
<tr>
<td>7</td>
<td>18:52</td>
<td>Over-the-Shoulder Close-Up</td>
<td>Salas blickt Caine an:</td>
<td>für dich sprechen.**</td>
<td>„Für dich aber auch nicht.”</td>
</tr>
<tr>
<td>9</td>
<td>19:11</td>
<td>Großer Aufnahme</td>
<td>Salas sieht leicht nach oben und blickt Caine an:</td>
<td>„Hamilton ist tot.“</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>19:17</td>
<td>Großer Aufnahme</td>
<td>Caine sieht sie nur wortlos an</td>
<td>„Ja.“</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>19:19</td>
<td>Großer Aufnahme</td>
<td>Salas erwidert den Blick und sagt:</td>
<td>„Und wahrscheinlich ist es das, was mir ein Überschreiten der Grenze unmöglich macht.“</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>19:25</td>
<td>Großer Aufnahme</td>
<td>Schnitt auf Caine</td>
<td>„Du kannst nichts für Raymonds Tod. Wir beide nicht. Du hast ihn mehr als jeder andere ständig gewarnt, vorsichtig …“</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Caine entgegnet:</td>
<td>… zu sein.“</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>„Ich hätte nachdrücklicher sein“</td>
<td></td>
</tr>
</tbody>
</table>
| Zeile | Uhrzeit | Sicht | Bild | Dialog | Register
|-------|---------|-------|------|--------|----------
| 13    | 19:28   | Normalsicht | Großenaufrnahme | Salas kontert: | „Das ist nicht wahr! Das stimmt doch nicht.“
| 14    | 19:32   | Normalsicht | Großenaufrnahme | Caine sieht sie an und seufzt lediglich: | „Hm.“
| 15    | 19:34   | Normalsicht | Großenaufrnahme | Salas sagt sanfter: | „Aber ich weiß, was du…“
| 16    | 19:37   | Normalsicht | Großenaufrnahme | Blick auf Caine | … meinst.“
| 17    | 19:40   | Normalsicht | Großenaufrnahme | Salas sieht Caine an und sagt: | „Okay…“
| 18    | 19:43   | Normalsicht | Großenaufrnahme | Caine blickt sie wortlos an | „Ich nehme an, ich muss die Dinge im Moment so nehmen, wie sie sind.“
| 20    | 19:51   | Normalsicht | Großenaufrnahme | Caine nickt kaum merklich und lächelt traurig | „Aber ich gehe nicht mit Statler aus. Dieses Grinsen…“
| 21    | 19:52   | Normalsicht | Over-the-Shoulder Close-Up | Salas spricht weiter, mit sarkastischem Unterton: | „Das ist mir auch aufgefallen.“
| 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: | „Aham…“
| 24    | 20:03   | Leichte Untersicht | Over-the-Shoulder Close-Up | Caine sieht Salas an und sagt: | „Ubrigens, jetzt…“
<table>
<thead>
<tr>
<th>Zeit</th>
<th>Typ</th>
<th>Einstellung</th>
<th>Szene</th>
<th>Dialogtext</th>
</tr>
</thead>
<tbody>
<tr>
<td>20:05</td>
<td>Normalsicht</td>
<td>Nahe</td>
<td>Salas blickt Caine an</td>
<td>„… tut er’s grade wieder.“</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Salas verdreht die Augen und lächelt</td>
</tr>
<tr>
<td>20:06</td>
<td>Leichte Untersicht</td>
<td>Over-the-Shoulder Close-Up</td>
<td>Caine geht nach hinten weg und sagt:</td>
<td>„Wir sehen uns.“</td>
</tr>
<tr>
<td>20:07</td>
<td>Leichte Untersicht</td>
<td>Amerikanische Einstellung</td>
<td>Salas erwidert:</td>
<td>„Ja, klar.“</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Salas bleibt stehen, Caine sagt im Hintergrund noch zu Restaurantangestellten:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>„Dankeschön, wiedersehen.“</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Salas fährt sich durch die Haare, die Rezeptionistin verabschiedet sich von Caine:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>„Schönen Abend noch.“</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Salas dreht sich nach links um und geht nach rechts aus dem Bild</td>
</tr>
</tbody>
</table>
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.


Bitte benutzen 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 Fragebogen und einen Multimedia-Teil. Für letzteren benötigen Sie unbedingt Sound - überprüfen Sie bitte vor dem Start der Befragung, dass dieser aktiviert und laut genug eingespielt 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, es ist auf Personen von 16 Jahren abwärts abgestimmt - ich freue mich über rege Beteiligung aller Altersklassen. Vielen Dank für Ihre Mithilfe!

Dankeschön und viel Vergnügen,
Bernadette Geißler
### Stimmungserhebung (Teil 1 von 6)

<table>
<thead>
<tr>
<th>Gefühl</th>
<th>Gar nicht</th>
<th>Wenig</th>
<th>Mittelmäßig</th>
<th>Ziemlich</th>
<th>Sehr stark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freude</td>
<td></td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Vertrauen</td>
<td></td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Angst</td>
<td></td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Überraschung</td>
<td></td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Traurigkeit</td>
<td></td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ekel</td>
<td></td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Wut</td>
<td></td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Neugierde</td>
<td></td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>
### Fragebogen (Teil 2 von 6)

**Phantasie**

<table>
<thead>
<tr>
<th>Frage</th>
<th>Trifft gar nicht zu</th>
<th>Trifft nicht zu</th>
<th>Trifft teilweise zu</th>
<th>Trifft zu</th>
<th>Trifft sehr zu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ich finde es etwas übertrieben, sich in Kinofilme oder Fernsehserien hineinzusteuern.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>In den Gefühlen von Romanfiguren lebe ich mich richtig binein.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Ich neige dazu, Kinofilme derart mitzusehen, dass ich empfinde, als wäre ich selbst eine der handelnden Personen.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Ich neige dazu, Fernsehserien derart mitzuerleben, dass ich empfinde, als wäre ich selbst eine der handelnden Personen.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Bei einem interessanten Roman stelle ich mir vor, wie ich in so einer Situation zurecht käme.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Es passiert mir eher selten, in einem guten Kinofilm besonders aufzugehen.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Es passiert mir eher selten, in einer guten Fernsehserie besonders aufzugehen.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Bei einer guten Fernsehserie stelle ich mir vor, wie ich wohl in so einer Situation zurecht käme.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

**Reale Situationen**

<table>
<thead>
<tr>
<th>Frage</th>
<th>Trifft gar nicht zu</th>
<th>Trifft nicht zu</th>
<th>Trifft teilweise zu</th>
<th>Trifft zu</th>
<th>Trifft sehr zu</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Reale Situationen

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

Weiter
### Serienkenntnis-Erhebung (Teil 3 von 6)

<table>
<thead>
<tr>
<th>Bild</th>
<th>Schätzen Sie, Wie oft haben Sie diese Fernsehserie (auf Deutsch) INSGESAMT bereits gesehen?</th>
<th>Noch nie</th>
<th>1 Mal</th>
<th>2 bis 5 Mal</th>
<th>5 bis 10 Mal</th>
<th>Öfter als 10 Mal</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="CSI Miami" /></td>
<td>CSI Miami</td>
<td>●</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td><img src="image" alt="Damages - Im Netz der Macht" /></td>
<td>Damages - Im Netz der Macht</td>
<td>●</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bild</th>
<th>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?</th>
<th>Verfolgte ich nicht regelmäßig</th>
<th>1 Mal pro Monat</th>
<th>2 bis 3 Mal pro Monat</th>
<th>1 Mal pro Woche</th>
<th>Öfter als 1 Mal pro Woche</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="CSI Miami" /></td>
<td>CSI Miami</td>
<td>●</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td><img src="image" alt="Damages - Im Netz der Macht" /></td>
<td>Damages - Im Netz der Macht</td>
<td>●</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>
Video-Fragebogen (Teil 4 von 6)

Bitte beachten Sie die Instruktionen:

- Überprüfen Sie, dass der Sound Ihres Computers aktiviert ist bzw. die Lautsprecherbogen Ihres Computers eingeschaltet sind.
- Warten Sie, bis die Videos geladen sind (kann einige Sekunden dauern).
- Wenden Sie sich nicht: Die Videosequenzen sind sehr kurz. Es gibt Videos mit Bild und Ton, Videos mit schwarzen Bild (nur Ton) sowie Videos ohne Ton (nur Bild).
- Sehen Sie sich ein Video an (wenn möchten auch mehrmals - ich empfehle jedoch nicht mehr als 3 Mal pro Video).
- Klicken Sie bitte nach jeder Gefühlseinschätzung auf "Speichern" und gehen zum nächsten Videoispiel über.

Beispiel 1 (nur Bild, kein Ton)

Was denken Sie, welches Gefühl vermittelt der/die Sprecherin/Schauspielerin?

- Freude
- Vertrauen
- Angst
- Erstaunen
- Traurigkeit
- Ekel/Absehen
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

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)
Heispiel 5 (Bild & Ton)

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

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

Speichern

Heispiel 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

Beispiel 7 (nur Ton, kein Bild)
Beispiel 7 (nur Ton, kein Bild)

Was denken Sie: Welches Gefühl vermittelt der/die Sprecherin/Schauspielerin?

- Freude
- Vertrauen
- Angst
- Erstaunen
- Tragik
- Ekel/Abscheu
- Wut
- Neugierde

Speichern
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).
- Wenden Sie sich nicht: Die Videosequenzen sind sehr kurz. Es gibt Videos mit Bild und Ton, Videos mit schwarzen Bild (nur Ton) sowie Videos ohne Ton (nur Bild).
- Seien Sie sich ein Video an (wenn Sie möchten auch mehrmals - ich empfehle jedoch nicht über die 3 Mal pro Video).
- 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
Beispiel 2 (nur Bild, kein Ton)

Was denken Sie: Welches Gefühl vermittelt der/die Sprecherin/Schauspielerin? 
- Freude
- Vertrauen
- Angst
- Erstaunen
- Traurigkeit
- Ekel/Absscheu
- Wut
- Neugierde

Beispiel 3 (Bild & Ton)
Beispiel 3 (Bild & Ton)

Was denken Sie. Welches Gefühl vermittelt die/das Sprecher/in/Schauspieler/in?

- Freude
- Vertrauen
- Angst
- Erstaunen
- Traurigkeit
- Ekel/Abneigung
- Wut
- Neugierde

Speichern

Beispiel 4 (nur Ton, kein Bild)
Beispiel 4 (nur Ton, kein Bild)

Was denken Sie. Welches Gefühl vermittelt der/die Sprecher/in/Schauspieler/in?

- Freude
- Vertrauen
- Angst
- Erstaunen
- Traurigkeit
- Ekke/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
- Eckel/Abneigung
- Wut
- Neugierde

Speichern

Beispiel 6 (Bild & Ton)
Beispiel 6 (Bild & Ton)

Was denken Sie. Welches Gefühl vermittelt der/die Sprecher/in/Schauspieler/in?

- Freude
- Vertrauen
- Angst
- Erstaunen
- Traurigkeit
- Ernst/Absehen
- Wut
- Neugierde

Beispiel 7 (nur Ton, kein Bild)
Beispiel 7 (nur Ton, kein Bild)

Was denken Sie: Welches Gefühl vermittelt der/die Sprecherin/Schauspielerin?

- Freude
- Vertrauen
- Angst
- Erstaunen
- Traurigkeit
- Ekel/Abstoßen
- 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).
- Wenden 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).
- Seien Sie sich ein Video an (wenn Sie möchten auch mehrmals - ich empfehle jedoch nicht über 3 Mal pro Video).
- 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/SchauspielderIn?

- Freude
- Vertrauen
- Angst
- Entsetzen
- Traurigkeit
- Ekel/Abneigung
Beispiel 2 (Bild & Ton)

Was denken Sie: Welches Gefühl vermittelt der/die Sprecherin/Schauspielerin?
- Freude
- Vertrauen
- Angst
- Erwartung
- Traurigkeit
- Ekel/Abscheu
- Wut
- Neugierde

Beispiel 3 (nur Bild, kein Ton)
Heispiel 3 (nur Bild, kein Ton)

Was denken Sie: Welches Gefühl vermittelt der/die Sprecherin/Schauspielerin?

☐ Freude
☐ Vertrauen
☐ Angst
☐ Erstaunen
☐ Traurigkeit
☐ Ekel/Abneigung
☐ Wut
☐ Neugierde

Speichern

Heispiel 4 (nur Ton, kein Bild)
Beispiel 4 (nur Ton, kein Bild)

Was denken Sie. Welches Gefühl vermittelt der/die Sprecher/in/Schauspieler/in?

- 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
- Enthobenheit
- Wut
- Neugierde

Speichern

Beispiel 7 (nur Ton, kein Bild)
Beispiel 7 (nur Ton, kein Bild)

Was denken Sie: Welches Gefühl vermittelt der/die Sprecherin/Schauspielerin?

- Freude
- Vertrauen
- Angst
- Erstarrung
- Traurigkeit
- Ekel/Abwertung
- Wut
- Neugierde

Speichern

Weiter
<table>
<thead>
<tr>
<th>Frage</th>
<th>Stimmt</th>
<th>Stimmt nicht</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ich gehe abends gerne aus</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ich habe einen Beruf, der mich voll befriedigt</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ich habe fast immer eine schlagfertige Antwort parat</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ich würde kaum zögern, auch alte und schwerbehinderte Menschen zu pflegen</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ich würde mich beim Kellner oder Geschäftsführer eines Restaurants beschweren, wenn ein schlechtes Essen serviert wird</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ich bin unternahmungslustiger als die meisten meiner Bekannten</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ich habe manchmal ein Gefühl der Teilnahmlosigkeit und inneren Leere</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Sind wir in ausgelassener Runde, so überkommt mich oft eine große Lust zu großen Streichen</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ich bin der Ansicht, die Menschen in den Entwicklungsländern sollten sich zuerst einmal selbst helfen</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ich bin traurig mit Menschen zusammen, die ich noch nicht kenne</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ich fühle mich über meine Familie hinaus für andere Menschen verantwortlich</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ich bin oft nervös, weil zu viel auf mich einstrahlt</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Wenn ich noch einmal geboren wurde, dann würde ich nicht anders leben wollen</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ich übernehme bei gemeinsamen Unternehmungen gerne die Führung</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ich lebe mit mir selbst im Frieden und ohne innere Konflikte</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ich finde, jeder Mensch soll sehen, wie er sich entscheidet</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ich habe gern mit Aufgaben zu tun, die schnelles Handeln verlangen</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ich denke oft, dass ich meinen Konsum einschränken müßte, um dann an benachteiligte Menschen abzugeben</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Meine Familie und meine Bekannten können mich im Grund eine kaum richtig verstehen</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ich kann in eine ziemlich langweilige Gesellschaft schnell Leben bringen</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Wenn jemand weint, möchte ich ihm an die Seele arbeiten und trösten</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Bei wichtigen Dingen bin ich bereit, mit anderen energisch zu konkurrieren</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Frage</td>
<td>Ja</td>
<td>Nein</td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
<td>----</td>
<td>------</td>
</tr>
<tr>
<td>Bei wichtigen Dingen bin ich bereit, mit anderen energisch zu konkurrieren</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ich fühle mich oft wie ein Pulverfass kurz vor der Explosion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ich würde mich eher als gesprächig bezecken</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Termindruck und Elektrizität lösen bei mir körperliche Beschwerden aus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In meinem bisherigen Leben habe ich kaum das verwundbaren können, was in mir steckt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ich bin ziemlich lebhaft</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ich mache mir oft Sorgen um meine Gesundheit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Es fällt mir schwer, vor einer großen Gruppe von Menschen zu sprechen oder vorzutragen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ich bin im Grunde eher ein ängstlicher Mensch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ich gebe gelegentlich Geld und Spenden für Katastrophenhilfe, Caritas oder andere wohltätige Organisationen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ich bin häufiger abgespannt, matt und erschöpft</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oft habe ich alles gründlich satt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ich schätze nur langsame Freundschaften</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manchmal habe ich ohne eigenen Grund ein Gefühl unbestimmter Gefahr oder Angst</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ich werde ziemlich leicht verlegen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wenn man mir eine kleine Geldspende gibt, ist nur das menschlich klingt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ich bin immer gut gelaunt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ich spiele anderen Leuten gerne einen harmlosen Streich</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ich errete leicht</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ich selten in beschränkter, unglücklicher Stimmung</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bei Geselligkeiten und öffentlichen Veranstaltungen bleibe ich lieber im Hintergrund</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ich träume oft von Dingen, die doch nicht verwirklicht werden können</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ich gründe viel über mein bisheriges Leben</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ich nehme mir viel Zeit, anderen Menschen geduldig zuhören, wenn sie von ihren Sorgen erzählen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oft rege ich mich zu rasch über jemanden auf</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statement</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Ich bin mit meinen gegenwärtigen Lebensbedingungen oft unzufrieden</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beim Reisen schaue ich lieber auf die Landschaft als mich mit den Mühselnden zu unterhalten</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Da der Staat schon für Sozialhilfe sorgt, brauche ich im Einzelnen nicht zu helfen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meine Laune wechselt ziemlich oft</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Es fällt mir schwer, den richtigen Gesprächstoff zu finden, wenn ich jemanden kennenlernen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alles in allem bin ich ausgesprochen zufrieden mit meinem bisherigen Leben</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ich habe oft das Gefühl, im Stress zu sein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meine Partnerbeziehung/Elle ist gut</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ich habe schon unentgelt beim Roten Kreuz, in anderen sozialen Einrichtungen oder in meiner Gemeinde geholfen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meistens blicke ich voller Zuversicht in die Zukunft</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Persönliche Daten (Teil 6 von 6)

<table>
<thead>
<tr>
<th>Fragestellung</th>
<th>Optionen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geschlecht</td>
<td>1. weiblich 2. männlich</td>
</tr>
<tr>
<td>Alter</td>
<td>Untermalchen</td>
</tr>
<tr>
<td>Höchster Schulabschluss</td>
<td>Pflichtschule, Lehre, Fachschule, Matura, Universität, Fachhochschule</td>
</tr>
<tr>
<td>Sind Sie berufstätig?</td>
<td>Ja, Ja und Studentin, Hausfrau/Hausmann oder sind Sie Schüler/in, Student/in</td>
</tr>
<tr>
<td>Familienstand</td>
<td>1. Ehefrau, Lebenspartner, verheiratet, getrennt lebend, geschieden, verwitwet</td>
</tr>
</tbody>
</table>
Vielen Dank für Ihre Teilnahme!

Fragen oder Anregungen bitte per Email an geissler_bernadette.
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)
Seit 06/2009 Redakteurin GamingXP (Computer- und Videospiele)