

Friedrich-Schiller-Universität  
Allgemeine Psychologie I  
Institut für Psychologie  
M. Sc. Christine Nussbaum, Prof. Dr. Stefan R. Schweinberger  
7. Fachsemester

**Untersuchung der systematischen Modulation des  
Adaptationseffektes bei der Wahrnehmung emotionaler  
Stimmen in Abhängigkeit des Sprechergeschlechts**

**Bachelorarbeit**

vorgelegt von:	Dorothea Berges dorothea.berges@uni-jena.de 180568
Studiengang:	Psychologie B. Sc. 7. Fachsemester
ErstgutachterIn:	M. Sc. Christine Nussbaum
ZweitgutachterIn:	Prof. Dr. Stefan R. Schweinberger
Ort und Datum:	Jena, 31.03.2022

# Inhaltsverzeichnis

Inhaltsverzeichnis.....	I
Abstract.....	1
1. Kommunikation und die Rolle des Sprechens.....	2
2. Theoretischer Hintergrund.....	3
2.1. Stimme.....	3
2.1.1 Stimme- und Sprachproduktion.....	3
2.1.2. Stimmenwahrnehmung.....	5
2.2. Emotionen.....	6
2.2.1. Definition von Emotion.....	6
2.2.2. Entstehung von Emotionen in der Stimme.....	7
2.2.3. Verarbeitung von Emotionen.....	9
2.3. Adaptation.....	12
2.3.1. Definition von Adaptation.....	12
2.3.2. Bisherige Ergebnisse.....	14
2.3.3. Einfluss des Sprechergeschlechts bei Adaptation.....	18
3. Experiment.....	20
3.1. Methode.....	20
3.1.1. ProbandInnen.....	20
3.1.2. Stimuli.....	21
3.1.2.1. Originale Stimulieraufnahmen.....	21
3.1.2.1. Target.....	22
3.1.2.1. Adaptor.....	22
3.1.3. Ablauf.....	22
3.1.4. Erhebung.....	25
3.2. Auswertung.....	25
3.2.1. Ergebnisse der Adaptationsblöcke.....	25
3.2.1.1. Adaptationsblock: Analyse des point of subjective equality (PSE).....	25
3.2.1.2. Adaptationsblock: Analyse der Standardabweichung.....	27
3.2.2. Ergebnisse des Baselineblocks.....	27
3.2.2.1. Baseline: Analyse des point of subjective equality (PSE).....	27
3.2.2.2. Baseline: Analyse der Standardabweichung.....	28

4. Diskussion.....	29
4.1. Limitationen der Studie.....	34
4.2. Weiterführende Forschung und folgende Fragestellungen.....	35
Abbildungsverzeichnis.....	III
Literaturverzeichnis.....	III
Anhang.....	VI

## **Abstract**

Obwohl Emotionen und die Fähigkeit diese zu erkennen grundlegend für die zwischenmenschliche Kommunikation sind, sind die Wahrnehmung von Emotionen und die Faktoren, die sie beeinflussen können, nicht ausreichend erforscht. Eine Möglichkeit Rückschlüsse auf die Wahrnehmung und Verarbeitung zu ziehen, ist der Adaptationseffekt. In emotionalen Gesichtern konnte mithilfe der Adaptation bereits gezeigt werden, dass Geschlecht und Emotionen nicht unabhängig verarbeitet werden und sich in der Wahrnehmung gegenseitig beeinflussen (Bestelmeyer et al., 2010a). Dies soll hier für emotionale Stimmen untersucht werden. In einem Online-Experiment wird eine zeitgleiche Adaptation auf die Emotionen Wut und Angst durchgeführt, wobei das Sprechergeschlecht (weiblich und männlich) der Stimuli systematisch variiert wurde. Es ergeben sich damit zwei Adaptationstypen (w\_ängstlich/m\_wütend und w\_wütend/m\_ängstlich), welche den ProbandInnen in einem within-subject-Design präsentiert werden. Damit wurde das Studiendesign von Bestelmeyer et al. (2010a) übernommen und von emotionalen Gesichtern auf emotionale Stimmen übertragen. In der Analyse zeigt sich eine Interaktion zwischen dem Geschlecht der Stimuli und der Wahrnehmung der Emotionen, sodass die Daten einen geschlechterspezifischen Nacheffekt der Adaptation aufweisen. Des Weiteren finden sich auch ohne Adaptation Unterschiede in der Wahrnehmung der Emotionen abhängig vom Sprechergeschlecht. Männer werden als wütender wahrgenommen, wohingegen Frauen als ängstlicher eingeordnet werden. Diese Unterschiede deuten entweder darauf hin, dass sich je nach Geschlecht die Ausdrucksfähigkeit für die Emotionen unterscheidet oder, dass Frauen und Männer in ihrer stereotypischen Rolle mit unterschiedlichen Emotionen verbunden werden. Die Daten des Experiments deuten auf eine Interaktion in der Verarbeitung von Emotion und Geschlecht hin.

## **1. Kommunikation und die Rolle des Sprechens**

Kommunikation ist alltäglich und vielfältig. Ein Nicken beim Vorbeigehen oder ein Lächeln kann genauso Teil der Kommunikation sein, wie ein langes Gespräch. Bei der Kommunikation spielt die menschliche Stimme eine zentrale Rolle, allerdings geht dies über die gesprochene Sprache hinaus (Watzlawick et al., 2017). Sprache enthält vielfältige Informationen über den Sprecher. Dazu gehören extralinguistische Signale, wie Identität, Geschlecht, Alter oder regionale Herkunft, aber auch paralinguistische Signale, wie Emotionen oder Gesundheitszustand des Sprechers (Schweinberger et al., 2014). Die Mechanismen der Sprachproduktion und die Vielfältigkeit der Informationen, die Sprache enthalten kann, sind eine wichtige Grundlage, um Kommunikation zu verstehen. Die Wahrnehmung dieser Signale ist dabei jedoch nicht absolut, sondern kann auf verschiedene Arten beeinflusst werden. Eine solche Einflussmöglichkeit ist die Adaptation. Dabei erzeugt eine einseitige Stimulation über längeren Zeitraum eine Wahrnehmungsverzerrung in die entgegengesetzte Richtung, den sogenannten kontrastiven Nacheffekt. Beispielsweise werden Stimmen auf einem Wut-Angst-Kontinuum als ängstlicher wahrgenommen, nachdem ProbandInnen vorher mehrere wütende Stimuli präsentiert wurden, bei Adaptation auf ängstliche Stimuli zeigt sich der umgekehrte Effekt (Bestelmeyer et al., 2010a). Diese Adaptationseffekte wurden für eine Reihe stimmlicher Signale gefunden, darunter Identität (Schweinberger et al., 2008), Alter (Zäske et al., 2013), Geschlecht (Webster et al., 2004), und Emotion (Bestelmeyer et al., 2010a). Die Adaptation erzeugt also einen Einfluss auf die Wahrnehmung eines über längere Zeit präsentierten Stimulus. Das Ziel dieser Bachelorarbeit ist die Untersuchung des Einflusses des Geschlechts bei der Adaptation bei emotionalen Stimmen. Inhaltlich werden zunächst die Grundlagen der Produktion und Wahrnehmung menschlicher Stimmen erläutert. Des Weiteren wird die Entstehung von Emotionen in der Stimme und die Verarbeitung von Emotionen dargelegt, sowie die Grundlagen zum Adaptationseffekt und dessen Einflüsse auf die Wahrnehmung von Reizen erläutert. In einem Experiment soll zudem untersucht werden, welche Rolle das Sprechergeschlecht bei der Verarbeitung emotionaler Stimmen hat und ob es den Effekt der Adaptation beeinflusst. Mit Hilfe des gleichen Adaptationsverfahrens wie bei Bestelmeyer et al. (2010a) wird die systematische Modulation des Adaptationseffektes bei der Wahrnehmung emotionaler Stimmen in Abhängigkeit des Sprechergeschlechts untersucht.

## **2. Theoretischer Hintergrund**

### **2.1. Stimme**

#### **2.1.1. Stimmen- und Sprachproduktion**

Um Kommunikation verstehen zu können, müssen zuerst die grundlegenden Vorgänge der Stimmenproduktion, sowie der Wahrnehmung der Stimme betrachtet werden.

Dabei gilt es zu beachten, dass die Begriffe Stimme und Sprache nicht als Synonyme verwendet werden können. Die in der Stimme enthaltenen Informationen umfassen mehr als die gesprochene Sprache (Kiese-Himmel, 2016). Auf körperlicher Ebene ist die Produktion von Sprache und Stimme zwar nicht zu unterscheiden, dennoch sollte beachtet werden, dass Stimme ein umfassenderer Begriff ist, welcher die Sprache miteinschließt. Die menschliche Stimme lässt sich über die Grundfrequenz F0 (entspricht der Tonhöhe), Intensität oder Lautstärke, Klangfarbe und zeitlicher Verlauf (entspricht der Sprechgeschwindigkeit) beschreiben (Banse & Scherer, 1996; Brück et al., 2011; Juslin & Laukka, 2003). Zusätzlich dazu spielen die Obertöne oder Formanten (F1-F8) eine Rolle. Die Grundfrequenzen bewegen sich in einem Bereich von 100 Hz bis 250 Hz, wobei der Hauptsprachbereich allgemein zwischen 250 und 4000 Hz liegt (Kiese-Himmel, 2016). Die Grundfrequenzen F0 sind verschieden für die Geschlechter. Männliche Stimmen bewegen sich in einem Bereich von 100 bis 150 Hz, wohingegen Frauen einen Frequenzbereich von 190 bis 250 Hz abdecken (Kiese-Himmel, 2016). Die Grundfrequenz ist durch diese deutlichen Unterschiede der wichtigste Parameter zur Erkennung des Sprechergeschlechts (Schweinberger et al., 2008; Skuk & Schweinberger, 2013).

Ein Modell zur Produktion von Stimme und eine Erklärung von Zusammenhängen der verschiedenen Frequenzen ist die Source-Filter-Theorie nach Fant (Fant, 1971). Die Theorie beschreibt Stimmenproduktion auf körperlichem Level und versucht die Komplexität der Sprachproduktion aufzuschlüsseln. Die Produktion der Stimme wird als ein mehrstufiger Prozess dargestellt, welcher in drei Teile *power*, *source* und *filter* gegliedert wird (Fant, 1971). Die *power* stellt den Ausgangspunkt dar. Dieser ist die Lunge, welche durch die Atmung den nötigen Luftstrom erzeugt, um einen Laut zu produzieren. Durch diesen Luftstrom werden in Folge die Stimmbänder in Bewegung gesetzt, sodass sie sich öffnen und schließen (Kiese-Himmel, 2016). Durch die Bewegung der Stimmbänder wird die Luft in Schwingung versetzt, sodass der Luftstrom in eine Druckwelle umgewandelt wird und ein hörbarer Ton entsteht. Dieser Ton hat die Grundfrequenz F0, welche sich aus der Öffnung und Schließung der

Stimm lippen pro Sekunde ergibt, was von Fant als *source* bezeichnet wird (Fant, 1971).

Durch die Veränderung des Luftstroms mit Hilfe gezielter Anpassung der Hohlräume im Mund- und Rachenbereich (*filter*) werden Vokale und Konsonanten gebildet (Fant, 1971). Das Signal, welches vom Kehlkopf durch die Stimmbänder gesendet wird, wird durch unterschiedliche Artikulatoren, wie Rachen, Zunge und Mundhöhle, angepasst (Fant, 1971). Damit kommt es zur Verstärkung beziehungsweise Abschwächung unterschiedlicher Frequenzen, sodass verschiedene Laute gebildet werden (Fant, 1971). Die Frequenzen, welche durch die Bewegung der Artikulatoren verändert - also verstärkt oder abgeschwächt werden - nennt man Formanten. Die Formanten unterscheiden sich nach ihren Frequenzen und ihre Zusammensetzung beschreibt die charakteristischen Frequenzen eines bestimmten Vokals (Fant, 1971; Kiese-Himmel, 2016). Für die Vokalqualität scheinen besonders die ersten vier (F1-F4) von Bedeutung zu sein, es gibt jedoch noch weitere, welche mitunter für die Wahrnehmung sozialer Signale in Stimmen von Bedeutung sein könnten, deren Einfluss auf die Stimme nicht genauer untersucht ist (Kreiman & Sidtis, 2011). Für die Formung von Vokalen sind allerdings vermutlich nur die F1-F4 von Bedeutung (Fant, 1971; Kiese-Himmel, 2016). Dass Menschen in der Lage sind, eine bekannte Stimme wieder zu erkennen, ist auf weitere mitschwingende Frequenzen zurückzuführen (Kiese-Himmel, 2016). Auch wenn die Stimme einer Person beim Sprechen unterschiedliche Höhen annehmen kann, so bleibt die charakteristische Klangfarbe der Stimme erhalten. Der Aufbau des Vokaltraktes sorgt dabei für einen individuellen Klang der Stimme (Kiese-Himmel, 2016). Dies kann man sich vorstellen wie bei einem Instrument. Der Resonanzkörper erzeugt hierbei eine charakteristische Klangfarbe, die Töne können sich allerdings voneinander unterscheiden. Der exakte Vorgang des Entstehens der Klangfarbe ist aufgrund der geringen Untersuchung der weiteren Formanten, neben den Formanten F1-F4, nicht genau bekannt. Es ist allerdings naheliegend, dass die Klangfarbe durch die Zusammensetzung aller Formanten entsteht (Kreiman & Sidtis, 2011).

Die Sprachproduktion wird ermöglicht durch schnelle und präzise aufeinander abgestimmte Bewegungen von Lunge, Stimmbänder und Artikulatoren, wie zum Beispiel die Zunge oder die Zähne. Diese Bewegungsabläufe werden durch spezifische Gehirnareale gesteuert. Bei der Sprachproduktion ist das Broca-Areal, welches im Frontalhirn der linken Hemisphäre im posterioren Bereich des Gyrus frontalis inferior (entspricht Brodmann Areal 44 und 45) lokalisiert ist (Schandry, 2016), das aktive Gehirnareal. Gefunden wurde es vom Chirurgen und Anthropologen Paul Broca, über

Beobachtungen von Patienten mit Sprachstörungen (Schandry, 2016). Er beobachtete, dass eine Schädigung dieses spezifischen Areals im Verlust der Produktionsfähigkeit von Lauten resultiert, jedoch nicht in Verlusten des Sprachverständnisses (Schandry, 2016). Dies brachte die Vermutung hervor, dass es gesonderte Areale für Stimmenproduktion und Stimmenwahrnehmung gibt.

### **2.1.2. Stimmenwahrnehmung**

Ein Gespräch ist ein wechselseitiger Prozess in der Kommunikation zwischen mehreren Personen. Um ein Gespräch zu führen, muss die Stimme gehört und deren Inhalt verarbeitet werden. Für diesen Prozess ist die Wahrnehmung der Stimme essenziell.

Die Wahrnehmung der Stimme lässt sich auf unterschiedlichen Ebenen betrachten: Höhen und Tiefen in der Stimme, charakteristische Eigenschaften der Stimme, wie Stottern oder Zittern, Geschwindigkeit oder die Klangfarbe (Schweinberger et al., 2014). Menschen beherrschen es, Unterschiede im Klang der Stimme wahrzunehmen (Schweinberger et al., 2014). Über diese Variabilität der Eigenschaften einer Stimme, und damit die gesamte auditorische Information einer Person, ist es möglich Stimmen, nach mehrfachem Hören, wiederzuerkennen (Schweinberger et al., 2014). Dabei wird nicht nur der Sprecher wieder erkannt. Auch beispielsweise Alter, Gesundheit, Attraktivität, soziale Herkunft oder Emotionen des Sprechers lassen sich auf Grundlage der Stimme ausmachen (Schweinberger et al., 2014). Es wird sogar spekuliert, ob es möglich ist, Persönlichkeitseigenschaften wie zum Beispiel Vertrauenswürdigkeit in Stimmen herauszuhören (Schweinberger et al., 2014).

Betrachtet man die Verarbeitung der Stimme auf neurologischer Ebene, so finden sich mehrere Areale, welche sowohl zur Sprachproduktion, als auch zur Wahrnehmung von Sprache und Stimme beitragen. Methodisch werden diese über bildgebende Verfahren, wie funktionelle Magnet-Resonanz-Tomographie (fMRT), untersucht. Wird ein Stimmenreiz aufgenommen, so lassen sich bilaterale Aktivitäten in den temporalen Stimmenarealen oder *temporal voice areas* (TVA) beobachten (Schweinberger et al., 2014). Diese Areale finden sich bilateral im medialen und anterioren Gyrus temporalis superior (Schweinberger et al., 2014). Die TVA zeigen bei der Präsentation von Stimmen eine stärker erhöhte Aktivität, als bei anderen auditiven Stimuli (Schweinberger et al., 2014). Dies ist ein Hinweis darauf, dass die Informationen aus einem Gespräch besonders in den TVA verarbeitet werden. Der Gyrus temporalis superior, auch als Wernicke-Areal bezeichnet, und weitere umliegende Areale werden



hingegen bei der Wahrnehmung von Sprache aktiviert. Auch das Broca-Areal, welches primär mit der Sprachproduktion in Verbindung gebracht wird, wird bei komplexen Aufgaben, wie beim Verstehen schwerer Texte, zusätzlich aktiviert (Schandry, 2016). Es zeigt sich, dass es für die Verarbeitung der Stimme eine Vielzahl von Arealen gibt, welche zum Teil abhängig von der transportierten Information aktiviert werden.

Zu den Informationen, die in der Stimme transportiert werden können, gehören unter anderem Informationen zur emotionalen Lage des Sprechers (Schweinberger et al., 2014). Was Emotionen sind und wie sie über die Stimme ausgedrückt werden können, wird im folgenden Abschnitt behandelt.

## **2.2. Emotionen**

### **2.2.1. Definition von Emotion**

Emotionen sind komplex und vielschichtig, sodass ein Erregungszustand des Gehirns, eine subjektive Empfindung oder eine physische Reaktion mit Emotionen beschrieben werden können (Schandry, 2016).

Paul Ekman teilt auf Grundlage von charakteristischen Merkmalen die Emotionen in Familien ein. Emotionen werden dabei über den Gesichtsausdruck, die Stimme, die automatische physiologische Reaktion und den Auslöser beschrieben (Ekman & Cordaro, 2011). Laut Ekman lassen sich auf Grundlage dieser Aspekte klassisch sechs Basisemotionen unterscheiden: Wut, Angst, Überraschung, Trauer, Ekel und Freude (Ekman & Oster, 1979). In späteren Publikationen findet sich die zusätzliche siebte Basisemotion Verachtung (Ekman & Cordaro, 2011). Die Basisemotionen bilden laut Ekman eine übergeordnete Kategorie, welche sich durch ihre gesamte Beschaffenheit von allen anderen Emotionen abgrenzen lassen. Diese Emotionen finden sich über alle Kulturen hinweg, was auf eine evolutionär bedingte Veranlagung hindeuten kann (Schandry, 2016). Ekman begründet den evolutionären Einfluss über die Reaktion auf emotionsauslösende Situationen, sodass sich in solchen Situationen jene bestimmten Emotionen beobachten lassen, welche sich evolutionär bewährt haben (Ekman & Cordaro, 2011). Besteht zum Beispiel Gefahr, so hat sich die Emotion Angst vor der Emotion Freude durchgesetzt.

Emotionen finden sich in unterschiedlichen Situationen und der Ausdruck von Emotionen kann die zwischenmenschliche Kommunikation beeinflussen. Ein fundamentaler Weg und möglicherweise der primäre Weg in der Kommunikation seinen Emotionen Ausdruck zu verleihen, ist über die Stimme (Darwin, 1872; Schweinberger et al., 2014).

### 2.2.2. Entstehung von Emotionen in der Stimme

Wie bereits erwähnt sind die Informationen, die über die Stimme transportiert werden können, vielfältig. Eine dieser Informationen ist die Emotion des Sprechers. Dabei ist es nicht nötig, dass der Gesprächspartner seine Gefühlslage mitteilt, sondern Hörer sind in der Lage, die Emotion aus dem Klang der Stimme zu erkennen (Schweinberger et al., 2014).

Banse und Scherer untersuchten 1996 die Zusammensetzung der Stimme bei verschiedenen Emotionen. Dabei wurden zwei Mechanismen unterschieden: das Encoding, also wie Emotionen überhaupt über die Stimme ausgedrückt werden, und das Decoding, was den Wahrnehmungsprozess der Emotionen aus der Stimme beschreibt (Banse & Scherer, 1996).

Das Encoding bezieht sich auf Emotionen und die Auswirkungen auf den Körper. Diese werden in der Prozessbeschreibung miteinander betrachtet, da Emotionen zu einer physischen Reaktion führen (Ekman, 2016; Ekman & Cordaro, 2011; Schandry, 2016), welche sich unter anderem in der Stimme beobachten lässt. Daraus lässt sich schließen, dass Emotionen die Stimmproduktion beeinflussen können. In Verbindung mit der Source-Filter Theorie, lässt sich feststellen, dass die Emotion Veränderungen sowohl in *source* und als auch *filter* bewirken können (Bachorowski, 1999).

Dies geschieht durch die akustische Manipulation des Luftstroms, welche den Ausdruck der Emotion bedingt (Bachorowski, 1999). Darüber können die Grundfrequenzen oder aber auch die Klangfarbe der Stimme angepasst werden, sodass es zu Modifikationen in der Stimmenproduktion kommt (Bachorowski, 1999). Ändert sich beispielsweise die Muskelspannung um die Lunge, so verändert sich der subglottale Druck, wodurch die Luft mit mehr Druck zu den Stimmlippen gepresst und die Grundfrequenz angepasst wird (Bachorowski, 1999). Des Weiteren ist zum Beispiel die Filterung des Luftstroms bei einer lächelnden Person anders, als bei einer Person, die nicht lächelt, wodurch eine andere Klangfarbe entsteht. Diese Varianz bedingt dabei eine Veränderung der Frequenzen, sowohl in der Grundfrequenz, als auch in den Formanten (Bachorowski, 1999).

Betrachtet man den Entstehungsprozess der Emotionen in der Stimme genauer, so finden sich zwei Möglichkeiten, Emotionen auszudrücken. Auf der einen Seite steht die nicht-sprachliche Vokalisation. Diese beschreibt einen Emotionsausdruck, der mit Lauten verbunden ist, allerdings nicht mit Sprache. Dazu zählen Lachen, Schreien oder Weinen (Scherer, 1986). Dem gegenüber stehen paralinguistische Signale, welche die gesprochene Sprache begleiten und auch als Sprachprosodie bezeichnet werden

(Scherer, 1986). Dazu zählen unter anderem Sprechpausen, Intonation, Tonhöhe oder Frequenz, Klangfarbe und Lautstärke, also emotionsgebundene Modifikationen verschiedener Sprachparameter. Es sind jene Signale, die im gesamten Stimmbild emotionale Informationen tragen können und somit die Interpretation ein und derselben Aussage völlig verändern können (Scherer, 1986).

Die paralinguistischen Signale in der Stimme können den Ausdruck der Emotionen über akustische Parameter beeinflussen. Aus dem Prozess des Encodings, wonach Emotionen durch akustische Signale kodiert werden, würde es naheliegen, dass es akustische Muster gibt, welche die verschiedenen Emotionen charakteristisch beschreiben (Bachorowski, 1999; Banse & Scherer, 1996; Scherer, 1986).

Eine Möglichkeit, die Kodierung von Emotionen zu veranschaulichen, ist die Komponenten-Prozess-Theorie (Banse & Scherer, 1996). Diese Theorie stellt eine Emotion als eine Reaktion des Körpers auf einen externen oder internen Stimulus dar. Die Bewertung des Stimulus sorgt dafür, dass die fünf Subsysteme Kognition, physiologische Regulation, Motivation, Bewegungsausdruck und die Fähigkeit der eigenen Kontrolle zusammenarbeiten (Banse & Scherer, 1996). Die sonst unabhängigen Systeme arbeiten dann parallel. Diese Zusammenarbeit drückt sich in den *stimulus evaluation checks* (SECs) aus (Banse & Scherer, 1996). Über mehrere Rückkopplungsschleifen werden für verschiedene Emotionen prototypische Profile der SECs aktiviert (Banse & Scherer, 1996; Scherer, 1986).

Auf Grundlage dieser Theorie machte Scherer Annahmen über die Zusammensetzung charakteristischer Akustikprofile für einzelne Emotionen (Banse & Scherer, 1996). Als Ausgangspunkt der Untersuchung dieser Muster der Emotionen dienen die Basisemotionen. Die Emotionen mit der höchsten Grundfrequenz sind Wut, panische Angst, Verzweiflung und Euphorie (Banse & Scherer, 1996). Verachtung und Langeweile werden mit einer niedrigen Grundfrequenz in Verbindung gebracht (Banse & Scherer, 1996). Des Weiteren wird Traurigkeit mit einer verlangsamten Stimme assoziiert; Verzweiflung, Euphorie und Wut mit einer erhöhten Sprechgeschwindigkeit (Banse & Scherer, 1996). Diese Annahmen zu den akustischen Mustern wurden in einer Meta-Analyse von Juslin und Laukka (2003) untersucht. Die Analyse ergab, dass die Annahmen Scheres auf Grundlage seiner Komponenten-Prozess-Theorie grundsätzlich mit der Empirie übereinstimmen (Juslin & Laukka, 2003).

Die Theorie unterstreicht die Wechselwirkung zwischen Emotionen und der körperlichen Reaktion und veranschaulicht, wie nicht-sprachliche und paralinguistische Signale Einfluss auf den emotionalen Ausdruck nehmen können. Sie

kann auch als Erklärungsgrundlage für die Veränderung des Grads der neuronalen Aktivierung (*arousal*) durch Emotionen dienen, denn eine erhöhte neuronale Aktivierung kann eine Veränderung im Emotionsausdruck hervorrufen (Bachorowski, 1999; Banse & Scherer, 1996).

Der Ausdruck von Emotionen kann jedoch abhängig von der Situation sehr variabel sein. Die äußeren Gegebenheiten können den Ausdruck von Emotionen je nach Situation und Sprecher beeinflussen (Brück et al., 2011). Jeder kann auf die gleiche Situation unterschiedlich reagieren, wonach es sehr komplex wäre, einen universellen Ausdruck von Emotionen zu erstellen.

Durch die Variabilität des Ausdrucks ist eine exakte Aufschlüsselung der Emotionen in der Stimme kompliziert. Meta-Analysen können Ansätze erstellen und den Versuch unternehmen, Einflussfaktoren aufzulisten und ihr Einwirken zu analysieren (Juslin & Laukka, 2003). Es zeigen sich hier Parallelen zu den theoretischen Vorhersagen von Scherer (1986) bezüglich der Zusammensetzung der akustischen Signale von Emotionen (Juslin & Laukka, 2003).

Der emotionale Ausdruck ist komplex. Ob die Emotion dabei durch die Umwelt, die Kognition oder durch eine spezifische Kombination, ausgelöst wird, bleibt weiter zu erforschen. Die Empirie legt nahe, dass es spezifische Parameter gibt, welche Emotionen in der Stimme charakteristisch beschreiben. Zwar gibt es noch keine genaue Evidenz über die konkreten stimmlichen Parameter aller Emotionen und wie genau sich diese zusammensetzen (Brück et al., 2011), aber dennoch scheint es Menschen erstaunlich gut möglich zu sein, diese zu erkennen.

### **2.2.3. Verarbeitung von Emotionen**

Der entgegengesetzte Prozess zum zuvor dargelegten Encoding ist das Decoding (Banse & Scherer, 1996). Decoding beschreibt den Wahrnehmungsprozess der Emotionen, in welchem die aufgenommenen Stimuli aufgeschlüsselt werden, sodass die in den stimmlichen Parametern enthaltenen Informationen verarbeitet werden können (Banse & Scherer, 1996).

Das Decoding beinhaltet damit auch die Verarbeitung jener Parameter, welche die Emotion in der Stimme beschreiben. Neurologische Evidenz weist darauf hin, dass die Verarbeitung von Emotionen vor allem in der rechten Hemisphäre stattfindet (Belin et al., 2011). Eine Beschränkung auf die rechte Gehirnhälfte wird allerdings der Komplexität von Emotionen nicht gerecht (Belin et al., 2011).

Die Frage, die sich bei der Verarbeitung der Emotionen stellt, ist, welche Parameter genau verwendet werden, um eine Emotion als solche zu erkennen (Brück et al., 2011). Frühe Theorien der Stimmenwahrnehmung beschreiben bei der Verarbeitung mehrerer Informationen parallele Verarbeitungsprozesse, welche interagieren können (Belin et al., 2011). Mittlerweile setzt sich die Sichtweise eines breiten Netzwerks mit multimodaler Verarbeitung der Stimuli durch (Brück et al., 2011; Young et al., 2020). Forschung zeigt, dass auf neurologischer Ebene bei der Verarbeitung von Emotionen komplexe neuronale Verbindungen eine Rolle spielen, was es kompliziert macht, den Prozess der Emotionsverarbeitung unabhängig von anderen Eigenschaften der Stimuli zu betrachten (Young et al., 2020). Zusammengefasst deutet die neurologische Forschung auf eine multimodale Verarbeitung hin, wobei das limbische System mit Thalamus und Amygdala, sowie der mediale superiore temporale Kortex mit TVA und dem primär auditiven Kortex bei der Verarbeitung von Emotionen aktiviert werden (Brück et al., 2011).

Die Verarbeitung von Emotionen lässt sich zum einen auf neurologischer Ebene betrachten, aber auch die Verhaltensforschung bringt interessante Ergebnisse, welche Rückschlüsse auf die neurologische Verarbeitung zulassen. Ähnlich, wie es Menschen möglich ist, unterschiedliche Identitäten anhand ihrer Stimme zu erkennen - erstmals nachgewiesen durch Adaptation auf verschiedene Identitäten durch Bestelmeyer und Mühl (2021) - können Menschen überzufällig die Emotionen in der Stimme korrekt zuordnen (Juslin & Laukka, 2003). Interessant ist hierbei, dass diese Fähigkeit abhängig von der zu kategorisierenden Emotion unterschiedlich stark ausgeprägt ist (Juslin & Laukka, 2003). Betrachtet man die Fehler von Versuchspersonen bei der Erfassung von Emotionen, so zeigen sich Verwechslungsmuster zwischen den Emotionen, das heißt, dass manche Emotionen häufiger verwechselt werden als andere (Banse & Scherer, 1996; Juslin & Laukka, 2003). Ausgehend von den Basisemotionen zeigt die Empirie, dass Traurigkeit und Freude selten verwechselt werden, wohingegen Traurigkeit und Zärtlichkeit häufiger nicht korrekt erkannt werden (Juslin & Laukka, 2003). Wichtig ist dabei zu bemerken, dass die Trefferquote der korrekten Erkennung steigt, wenn die Entscheidung auf einer umfangreicheren Menge an Informationen beruht. Dies zeigt sich darin, dass die korrekte Einschätzung einer Emotion auf Basis der Stimme eine höhere Fehlerquote zeigt, als bei der gemeinsamen Präsentation von Stimme und Gesicht (Belin et al., 2011). Dieses Wiedererkennen der Emotionen ist grundlegend für die zwischenmenschliche Kommunikation und die Interpretation von Situationen (Juslin & Laukka, 2003). Juslin und Laukka (2003) beschreiben das

korrekte Erfassen von Emotionen als Grundlage für weiteres Agieren in beispielsweise Gefahrensituationen.

Obwohl alle Menschen Emotionen überzufällig gut erkennen können (Juslin & Laukka, 2003), so gibt es dennoch substanzielle Unterschiede in der Fähigkeit, Emotionen in Stimmen korrekt zu klassifizieren. Musiker können Emotionen beispielsweise besser erkennen und zeigen so eine höhere Trefferquote als eine Kontrollgruppe (Lima & Castro, 2011). Personen mit Autismus hingegen zeigen unterschiedliche Werte je nach Emotionen. Eine Meta-Analyse ergab, dass Personen mit Autismus Angst deutlich schlechter erfassen, als beispielsweise Freude (Uljarevic & Hamilton, 2013).

Es gibt diverse Forschung über die Zusammensetzung der Areale, welche die Verarbeitung der Stimme beeinflussen und wie genau der Verarbeitungsprozess aufgebaut ist. Hierbei stehen sich bildgebende Verfahren und die Beschreibung der Wahrnehmung von Emotionen in einem stufenweisen Prozess und der Versuch der akustischen Aufschlüsselung von Emotionen gegenüber. Für Menschen scheint das Differenzieren verschiedener Emotionen relativ intuitiv zu sein, aber Forschern ist es bisher noch nicht möglich, eine konkrete und reliable Aufschlüsselung der Emotionen in der Stimme zu erstellen (Scherer, 1986). Eine insgesamt Betrachtung des emotionalen Erfassungsprozesses bleibt ein Ziel für weitere Forschung.

Trotz der umfassenden Untersuchung der neuronalen Verarbeitung von Emotionen sind die genauen Vorgänge nicht vollends erforscht. Es finden sich Hinweise darauf, dass die Emotion des Betrachters bei der Wahrnehmung von emotionalem Material einen Einfluss auf die Verarbeitung der präsentierten Emotionen hat (Banse & Scherer, 1996). Das heißt, dass die Wahrnehmung von Emotionen relativ ist zur bereits vorhandenen Information (Banse & Scherer, 1996).

Ein Paradigma, in dem sich die Relativität der Wahrnehmung einer Emotion veranschaulichen lässt, ist die Adaptation.

## **2.3. Adaptation**

### **2.3.1. Definition von Adaptation**

Beim Effekt der Adaptation handelt es sich um einen Wahrnehmungseffekt. Der Prozess der Adaptation resultiert in einer veränderten Wahrnehmung eines Stimulus nach konstanter Präsentation in die entgegengesetzte Richtung auf einem Kontinuum der Parameter des ursprünglichen Reizes. Dies führt zum sogenannten kontrastiven Nacheffekt, sodass ein ambiguer Reiz eher als das Gegenteil des zuvor präsentierten Reizes wahrgenommen wird (Bestelmeyer et al., 2010b). Zur Veranschaulichung ein Beispiel aus der Farbwahrnehmung: Man betrachtet das Farbkontinuum gelb-blau. Setzt man das Auge über längere Zeit einem gelben Reiz aus, beispielsweise über gelb getönte Gläser einer Sportbrille, und betrachtet nach dem Abnehmen der Brille eine weiße Wand, so wirkt diese blau (Bäuml, 1991). Die Wahrnehmung eines ambigen oder uneindeutigen Reizes wird durch die Präsentation vom gelben Reiz beeinflusst in die entgegengesetzte Richtung des Kontinuums, also zu blau (Bäuml, 1991). Man kann demnach durch Adaptation einen Nacheffekt in der Wahrnehmung erzeugen, welcher sich durch die veränderte Wahrnehmung uneindeutiger Reize weg vom zuvor präsentierten Reiz äußert.

Die Adaptation ist allerdings nicht der einzige Effekt, welcher die Wahrnehmung beeinflussen kann, sodass die Notwendigkeit besteht, die Adaptation von anderen Veränderungen der Wahrnehmung abzugrenzen. Wichtig ist die Unterscheidung zum Priming, welches einen gegenteiligen Effekt zu Adaptation hervorruft. Priming sorgt dafür, dass durch vorher präsentierte Information die Wahrnehmung nachfolgender Stimuli in Richtung der eben genannten vorherigen Reize manipuliert wird (Bestelmeyer et al., 2010b).

In einem klassischen Adaptationsparadigma nutzt man Stimuli, welche auf einem Kontinuum zwischen zwei Endpunkten liegen. Beispielsweise ein ängstlicher Gesichtsausdruck auf der einen Seite und auf der anderen Seite ein Gesichtsausdruck, welcher Wut ausdrückt (Bestelmeyer et al., 2010a). Zudem benötigt man eine ambigue Mitte, welcher keiner der beiden Gruppen eindeutig zugeordnet werden kann. In Experimenten wird dabei häufig mit Morphleveln gearbeitet, die das Kontinuum der präsentierten Stimuli in regelmäßigen Schritten beschreiben.

Um im Standardparadigma der Adaptation einen Effekt messen zu können, werden die Versuchspersonen in einem Vortest aufgefordert, die Stimuli auf dem besagten Kontinuum einem der beiden Kategorien des Kontinuums, im Beispiel wütend oder ängstlich, zuzuordnen. Diese Stimuli, welche durch die Versuchspersonen eingeschätzt

werden sollen, bezeichnet man als *Targets*. Eine erste Bewertung der *Targets* beschreibt den Ausgangspunkt der Wahrnehmung, durch welchen man in der Auswertung feststellen kann, ob es zu einer Veränderung der Wahrnehmung gekommen ist. In der Adaptationsphase werden den Versuchspersonen die *Adaptoren* präsentiert. Die *Adaptoren* befinden sich am Ende des festgelegten Kontinuums, wobei pro Adaptation nur eine Seite des Kontinuums gezeigt wird, sodass die ProbandInnen zum Beispiel nur wütende Gesichter sehen (Bestelmeyer et al., 2010b). Danach werden die Versuchspersonen aufgefordert, die *Targets* erneut zu bewerten, um eine mögliche Beeinflussung der Wahrnehmung im Vergleich zum Vortest zu messen. Bei erfolgreicher Adaptation lässt sich erkennen, dass die Bewertung der *Targets* sich abhängig von der durchgeführten Adaptation verändert hat (Bestelmeyer et al., 2010b).

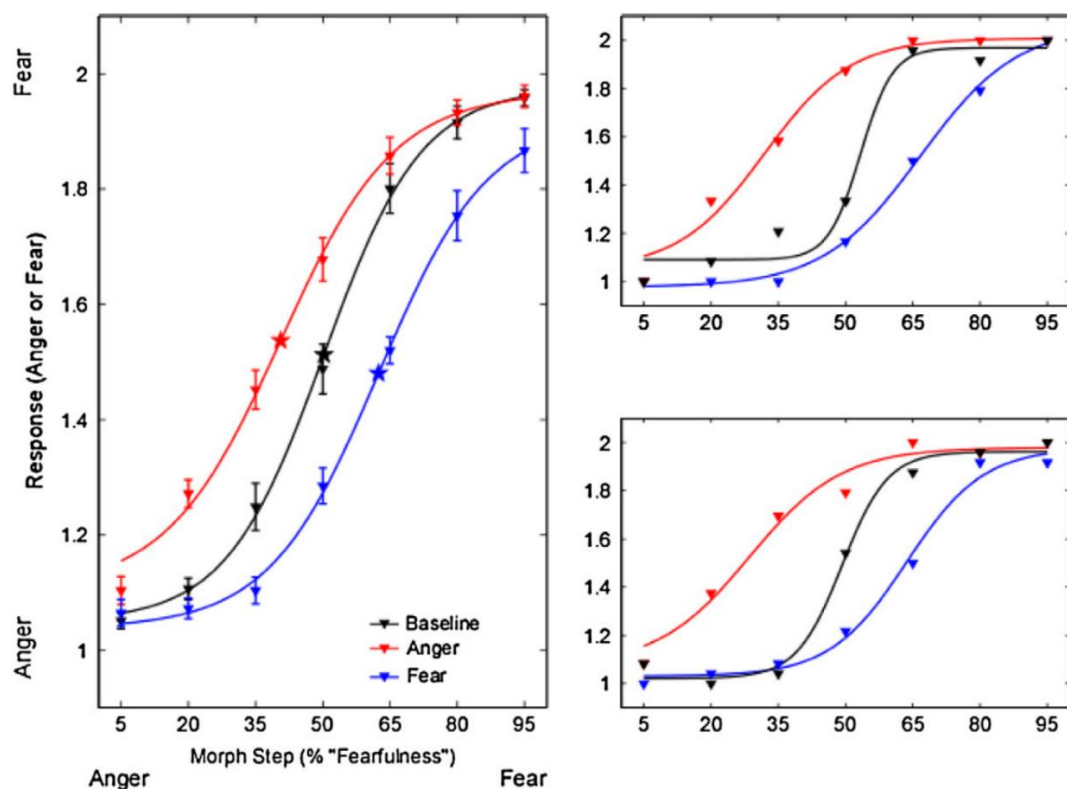


Abbildung 1. Darstellung des Adaptationseffekts bei Gesichtsausdrücken. Links: durchschnittlicher Effekt aller Versuchspersonen. Rechts: Ergebnisse individueller TeilnehmerInnen.  
Quelle: Bestelmeyer et al. (2010b), S. 220.

Abbildung 1 veranschaulicht die Ergebnisse eines klassischen Adaptationsexperiments. Der schwarze Graph stellt die Wahrnehmung der Gesichtsausdrücke dar ohne vorangegangene Adaptation. Der rote und blaue Graph veranschaulicht jeweils die Verschiebung des uneindeutigen Punktes in die entgegengesetzte Richtung des Ausgangsstimulus.



Das bedeutet, wenn die Versuchspersonen in der Adaptationsphase wütende Gesichter gesehen haben, so nehmen sie im Verlauf des Kontinuums von wütend nach ängstlich die Gesichter schneller als ängstlich wahr, als vor dem Adaptationsprozess (Bestelmeyer et al., 2010b). Der so genannte *point of subjective equality* (PSE) verschiebt sich. Der PSE bezeichnet jenen Punkt im Kontinuum, welcher von den ProbandInnen zu keiner der beiden Seiten eindeutig zugeordnet werden kann. Er befindet sich daher immer innerhalb des betrachteten Kontinuums. Ohne Adaptation befindet sich der PSE ungefähr bei dem 50% Morphlevel (Bestelmeyer et al., 2010b). Nach der Adaptation befindet sich dieser nicht mehr bei 50% – wie in Abbildung 1 sichtbar. Der PSE liegt für eine Adaptation auf Angst nun bei circa 40 % und bei einer Adaptation auf Furcht bei circa 65 % (Bestelmeyer et al., 2010b).

Mit Hilfe des Adaptationsparadigmas können Routen der Informationsverarbeitung untersucht werden und somit der Ablauf der Verarbeitung verschiedener Signale aufgeschlüsselt werden (Bestelmeyer et al., 2010b). Werden beispielsweise mehrere Informationen aus der Stimme präsentiert, wie das Alter und die Emotion, um zu untersuchen, ob eine altersspezifische Adaptation möglich ist, so kann auf die Emotion adaptiert werden und das Alter zusätzlich präsentiert werden (Zäske et al., 2013). Kommt es zu einem altersspezifischen Adaptationseffekt, können Annahmen über die neuronalen Pfade der Wahrnehmung aufgestellt werden (Zäske et al., 2013). Findet man einen spezifischen Nacheffekt nach der Adaptation, so lässt sich rückschließen, dass die präsentierten Informationen auf einem interaktionalen Pfad verarbeitet werden - ist dem nicht so, so werden die Informationen unabhängig voneinander verarbeitet (Zäske et al., 2013). Dies ist unter anderem für die Untersuchung der Wahrnehmung von Stimmen relevant (Belin et al., 2011; Webster et al., 2004).

Adaptation gibt Aufschlüsse über mögliche Routen der Informationsverarbeitung, welche nicht allein über die Beobachtung neuronaler Verknüpfungen erklärt werden können. Das Paradigma gibt die Möglichkeit, die neuronale Integration von Informationen aufzudecken und Netzwerke der Wahrnehmung zu erforschen.

### **2.3.2. Bisherige Ergebnisse**

Die biologische Psychologie kennzeichnet den Beginn der Forschung zur Adaptation. Phänomene, wie die Dunkeladaptation, Farbadaptation (Bäumel, 1991) oder die Adaptation zu Gerüchen und Geschmack bilden die Hintergründe der Adaptationsforschung (Schandry, 2016).

Betrachtet man beispielweise für einige Zeit ein bewegtes Bild, wie einen Wasserfall oder eine sich drehende Spirale und wechselt dann den Stimulus in ein stationäres, sich nicht bewegendes Bild, so ergibt sich eine scheinbare Bewegung beim Betrachten des stationären Bildes in die entgegengesetzte Richtung des ursprünglichen Stimulus (Bülthoff, H. H., Fahle, M., Gegenfurtner, K. R., Mallot H.A., 1998; Heller, D., Krummenacher, J., Bösch, U., 1998). Dieser Bewegungseffekt ist durch den Effekt der Adaptation erklärbar. Wird man einem Stimulus über längere Zeit ausgesetzt, so verschiebt sich die Wahrnehmung eines uneindeutigen Stimulus in Richtung des zuvor präsentierten Reizes.

Adaptationseffekte finden sich beispielweise auch bei Gesichtern, wie bei Webster et al. (2004), die den Adaptationseffekt bei Gesichtszügen nachweisen konnten. Der Adaptationseffekt konnte in den Dimensionen Geschlecht (männlich versus weiblich), ethnische Herkunft (kaukasisch versus asiatisch) und Gesichtsausdruck (Ekel versus Überraschung) gemessen werden (Webster et al., 2004). Dabei beziehen sich die Ergebnisse für Geschlecht und ethnische Herkunft ausschließlich auf die natürlichen Parameter des Gesichts ohne Ausdruck, der Einfluss der Emotionen wurde hier nicht untersucht (Webster et al., 2004).

Auch im Bereich der Emotionen bezieht sich die Adaptationsforschung in früheren Phasen auf die Wahrnehmung von Emotionen in Gesichtern. Dazu wurden Bilder von charakteristischer Mimik für verschiedene Emotionen manipuliert und mit anderen Emotionen gemorpht (Bestelmeyer et al., 2010a; Skuk & Schweinberger, 2013).

Eine Studie, welche sich mit der Untersuchung der Verarbeitung von emotionalen Gesichtern beschäftigt, ist die Studie von Bestelmeyer et al. (2010a). Hierbei wurde in zwei Experimenten mit Hilfe des Adaptationseffektes untersucht, ob die Verarbeitung emotionaler Gesichter geschlechterspezifisch und rassenspezifisch stattfindet (Bestelmeyer et al., 2010a). In einem ersten Experiment wurde auf die Emotionen Wut und Angst adaptiert, wobei zeitgleich auf beide Emotionen adaptiert wurde - nur für unterschiedliche Geschlechter (Bestelmeyer et al., 2010a). So wurden beispielsweise die Frauen wütend und die Männer ängstlich zur Adaptation präsentiert (Bestelmeyer et al., 2010a). Danach beurteilten die ProbandInnen ein Kontinuum zwischen den Emotionen Angst und Wut (Bestelmeyer et al., 2010a). Es wurde also überprüft, ob sich die Adaptationseffekte abhängig vom Geschlecht unterscheiden, um Rückschlüsse auf die Verarbeitung von Geschlechtern und emotionalen Gesichtern zu ziehen. Mit Hilfe des Adaptationseffekts konnte gezeigt werden, dass die Verarbeitung von Emotionen abhängig vom Geschlecht stattfindet, da sich die Adaptationseffekte der

Emotion nur für das jeweils präsentierte Geschlecht zeigten, aber nicht für das andere Geschlecht (Bestelmeyer et al., 2010a). Dies legt nahe, dass das Geschlecht beim Betrachten eines emotionalen Gesichtsausdruckes ebenfalls verarbeitet wird. Im zweiten Experiment wurde dasselbe Prinzip angewendet, nur in Bezug auf Rassen (Bestelmeyer et al., 2010b). Hierbei wurden die Versuchspersonen auf emotionale Gesichter mit afrikanischer und ost-asiatischer Herkunft adaptiert (Bestelmeyer et al., 2010a). Auch hier zeigte sich eine spezifische Adaptation abhängig von der Rasse (Bestelmeyer et al., 2010a).

Mit der Weiterentwicklung der Morphteknik wurde es möglich, nicht nur Gesichter, sondern auch Stimmen zu bearbeiten, womit es möglich wurde, den Adaptationseffekt nun auch bei der Stimme zu untersuchen (Schweinberger et al., 2008). Die erste Adaptation emotionaler Stimmen konnte von Bestelmeyer et al. (2010b) in „*Auditory adaptation in vocal affect perception*“ nachgewiesen werden. Des Weiteren konnte gezeigt werden, dass die Adaptationseffekte bei Stimmkarikaturen, also Stimmen, in denen vokale Parameter künstlich verstärkt werden, nicht größer ist als bei normal emotionalen Stimmen (Bestelmeyer et al., 2010b). Dies ist ein Hinweis darauf, dass die Verarbeitung emotionaler Stimmen auf einem erhöhten neurologischen Niveau stattfindet, da ein verstärkter Emotionsausdruck nicht direkt mit einem stärkeren Effekt einher geht (Bestelmeyer et al., 2010b).

Ein Einfluss des Geschlechts auf den Adaptationseffekt bei Emotionen konnte bereits in Gesichtern nachgewiesen werden (Bestelmeyer et al., 2010a). Studien zeigen, dass ein wichtiger Faktor, um Geschlechterunterschiede in der Stimme wahrzunehmen, die Grundfrequenz F0 ist (Schweinberger et al., 2014). Mit der Adoleszenz verändern sich die Frequenzen der Stimme und der sexuelle Dimorphismus, also die deutliche Unterscheidung der Geschlechter, stellt sich ein (Schweinberger et al., 2014). Die Geschlechter können nun besser auf Grundlage der Grundfrequenzen unterschieden werden (Schweinberger et al., 2014). Die Grundfrequenz allein reicht allerdings nicht aus, um die Adaptation auf das Geschlecht zu erklären (Schweinberger et al., 2008).

Zäske et al. (2013) untersuchten den Einfluss von Alter und Geschlecht auf die Adaptation und legten in ihrer Studie dar, dass die Wahrnehmung von Alter und Geschlecht miteinander zusammenhängen. In einem ersten Experiment wurde auf die Parameter jung versus alt adaptiert und in einem zweiten auf weiblich versus männlich. Besonders interessant ist, dass bei der Variation des Geschlechts zwischen Adaptationsdurchgang und Nachtest der Effekt der Adaptation sinkt. Wird mit einer männlichen Stimme die Adaptation im Adaptationsblock durchgeführt und die

Beurteilung der Stimuli im Nachtest in Folge mit einer weiblichen Stimme, so verringert sich der Effekt der Adaptation (Zäske et al., 2013). Dies lässt auf interaktionale Verarbeitung in der Wahrnehmung von Alter und Geschlecht schließen (Zäske et al., 2013). Ähnliche Ergebnisse im Zusammenhang zwischen der Verarbeitung emotionaler Stimmen und Geschlecht finden sich in Nussbaum et al. (2022). Hierbei wurde die Rolle der Grundfrequenz F0 und der Klangfarbe für die Adaptation auf Emotionen untersucht. In zwei Experimenten wurden die Rollen der Klangfarbe und der Grundfrequenz für die Adaptation auf Emotionen untersucht, mit dem Ziel herauszufinden, welche akustischen Parameter für die Adaptation von Bedeutung sind (Nussbaum et al., 2022). Auf Grundlage der Theorie, dass unterschiedliche akustische Parameter für die Entstehung von Emotionen in der Stimme von Bedeutung sind, vor allem jene, die mit der F0 zusammenhängen, wurde über Morphtechneik entweder die F0 oder die Klangfarbe der Stimuli modifiziert, wobei die anderen stimmlichen Parameter konstant gehalten wurden (Nussbaum et al., 2022). Die Ergebnisse zeigen größere Adaptationseffekte bei der Klangfarbe als bei der F0 (Nussbaum et al., 2022). In einem zweiten Experiment wurde das Sprechergeschlecht zwischen Adaptation und Bewertung der Stimuli variiert, was zur Folge hatte, dass der Adaptationseffekt nicht mehr signifikant war (Nussbaum et al., 2022). Es wurde gezeigt, dass die Klangfarbe für die Adaptation auf Emotionen eine wichtigere Rolle spielt, als die Grundfrequenz F0 und, dass durch Variation der Geschlechter der Adaptationseffekt unterdrückt wird (Nussbaum et al., 2022).

Die Adaptationsforschung stellt sich weiterhin die Aufgabe, unterschiedliche Verarbeitungsmuster zu analysieren. Mithilfe von immer neuen Experimenten zur Untersuchung verschiedener Einflüsse konnte erst kürzlich der Zusammenhang zwischen Stimmendiskrimination und Adaptation auf Identitäten in der Stimme bewiesen werden (Bestelmeyer & Mühl, 2021).

Die beschriebenen Experimente sollen einen Einblick in die bisherigen Ergebnisse im Feld der Adaptationsforschung geben. Es ist zu erwarten, dass es in Zukunft noch diverse weiterführende Forschung geben wird, vor allem zu Adaptationseffekten in der Stimme, da dies ein relativ gesehen junges Forschungsfeld ist und noch viele Möglichkeiten bietet.

### 2.3.3. Der Einfluss des Sprechergeschlechts bei Adaptation

Geschlechterspezifische Nacheffekte durch Adaptation auf die Wahrnehmung von Gesichtern konnten bereits nachgewiesen werden. Dies unterstützt die Ergebnisse von Zäske et al. (2013) und Nussbaum et al. (2022). Wie bereits erwähnt, wurde hierbei das Geschlecht der präsentierten Stimuli im Adaptationsteil und Targetteil variiert. Waren die *Adaptoren* weiblich, so wurden männliche *Targets* präsentiert und andersherum. Es zeigt sich, dass sich bei der Variation des Geschlechtes der Gesichter zwischen den Blöcken der Erhebung eine Verkleinerung des Adaptationseffekts messen lässt (Nussbaum et al., 2022; Zäske et al., 2013).

Weiterführende Forschung beschäftigt sich mit der Adaptation bei emotionalen Stimmen (Bestelmeyer et al., 2010b), sowie der Untersuchung der Adaptation bei Emotionen und Gesichtern in Zusammenhang mit beispielsweise dem Geschlecht (Bestelmeyer et al., 2010b; Bestelmeyer et al., 2010a).

Trotz der breiten Forschungsmöglichkeiten ist der Einfluss des Geschlechts des Sprechers auf den Adaptationseffekt bei emotionalen Stimmen wenig untersucht worden. Dies ist wohl darauf zurückzuführen, dass die Forschung zu emotionalen Stimmen und Emotionen in der Stimme selbst relativ jung ist. Um die Effekte der Adaptation messen zu können, benötigt man ein Kontinuum an Stimuli, welche ineinander überführt oder gemorpht werden können (Schweinberger et al., 2008). Diese Technik des Morphing ist für Stimmen weniger lange möglich, als das Morphing von Gesichtern, es bildet aber eine Voraussetzung für die Adaptationsforschung (Schweinberger et al., 2008).

Auf Grundlage der bisherigen Forschung ist das Thema dieser Arbeit die Untersuchung der Einflüsse des Sprechergeschlechts der Adaptationsstimuli auf den Adaptationseffekt bei emotionalen Stimmen. Dazu wird die systematische Modulation des Adaptationseffektes bei der Wahrnehmung emotionaler Stimmen in Abhängigkeit des Sprechergeschlechts untersucht. Dies soll in einem Online-Experiment untersucht werden. Dabei wird das gleiche Adaptationsverfahren wie Bestelmeyer et al. (2010a) genutzt. Allerdings wird die Modalität verändert, anstelle von emotionalen Gesichtern werden emotionale Stimmen verwendet. Es wird sich auf das Kontinuum der Emotionen Wut bis Angst bezogen, da die Empirie zeigt, dass sich diese Emotionen am besten voneinander unterscheiden lassen (Bestelmeyer et al., 2010b).

Die Theorie hinter dem für diese Arbeit durchgeführten Experiment bezieht sich auf die zeitgleiche Adaptation auf zwei Emotionen je am Ende eines Kontinuums (in diesem Fall wütend-ängstlich), wobei die Geschlechter systematisch variiert werden.

Das Ziel der Studie ist zu untersuchen, ob bei der Wahrnehmung von Emotionen über stimmliche Modalitäten geschlechterspezifisch adaptiert wird. Da bei einer zeitgleichen Adaptation die gleiche Anzahl an wütenden und ängstlichen Stimuli präsentiert wird, würden, bei einer unabhängigen Verarbeitung, die Stimuli sich gegenseitig herausmitteln. Damit käme es bei unabhängiger Verarbeitung zu keinem Adaptationseffekt. Werden Geschlecht und Stimme allerdings abhängig voneinander verarbeitet, so müsste es zu unterschiedlicher Wahrnehmung der Emotionen kommen, welche abhängig vom Geschlecht des Stimulus ist. Infolgedessen würde eine geschlechterspezifische Adaptation entstehen. Es wird die Hypothese dargelegt, dass es zu einer geschlechterspezifischen Adaptation kommen wird, was auf eine gemeinsame Verarbeitung der Informationen Emotion und Geschlecht aus der Stimme hindeuten würde. Dies bedeutet, dass sich die Wahrnehmung der Emotionen nach einer Adaptation auf die Emotion abhängig vom systematisch präsentierten, allerdings nicht adaptierten, Sprecher-geschlecht ändert.

### **3. Experiment**

Zur Untersuchung der Rolle des Sprechergeschlechts auf den Adaptationseffekt wurde ein Online-Experiment (Bestelmeyer et al. 2010a) durchgeführt. Das Experiment übernimmt den Aufbau der Studie von Bestelmeyer et al. (2010a). Das Ziel der Studie ist, zu zeigen, dass bei der Wahrnehmung von Emotionen in der Stimme geschlechterspezifisch adaptiert wird. Dazu wurde der Ablauf des Experiments von Bestelmeyer et al. (2010a) übernommen und auf die Modalität der Stimme übertragen. Ein geschlechterspezifischer Adaptationseffekt bei der Wahrnehmung von Emotionen wurde bereits bei Gesichtern nachgewiesen (Bestelmeyer et al. 2010a). Unser Ziel ist es diesen Effekt bei der Wahrnehmung von emotionalen Stimmen zu replizieren.

#### **3.1. Methode**

##### **3.1.1. ProbandInnen**

Im Vorfeld zu der Erhebung wurde eine Poweranalyse mit dem Paket „Superpower“ (Caldwell et al., 2019) durchgeführt mit einer mittleren Effektstärke  $f = .167$ , einem  $\alpha$ -Niveau = .05 und einer Power = 80 für die Interaktion zwischen Sprechergeschlecht und Adaptationstyp auf die Bewertung der Stimuli durchgeführt. Laut Analyse wird dazu eine ProbandInnenanzahl von  $n \geq 38$  benötigt. In der Ausschreibung der Studie wurde nach ProbandInnen zwischen 18 und 40 Jahren, Muttersprache Deutsch und einem gesunden Hörvermögen gesucht. Insgesamt wurde der Link zur Studie 56 mal angeklickt, wobei 40 vollständige Datensätze entstanden sind.

Nicht vollständige Datensätze wurden in der weiteren Analyse nicht berücksichtigt. Von den verbleibenden 40 Datensätzen wurden 7 von der Analyse ausgeschlossen, da sie bei der Schätzung der Parameter der kumulativen Gaussfunktion entweder einen zu schlechten Modellfit hatten ( $R^2 < .70$ ) oder einen PSE von unter 20 oder über 80, was mit den präsentierten Morphleveln nicht vereinbar ist. Zur Erinnerung: Der PSE ist jener Punkt, an welchem die ProbandInnen einen Stimulus nicht eindeutig einer Seite des Kontinuums zuordnen können und befindet sich daher innerhalb des Morphkontinuums.

Entsprechend gingen 33 Datensätze in die Analyse ein, mit einer Altersverteilung zwischen 19 und 39 Jahren (Alter: Mittelwert=22.97; Standardabweichung=3.65). Keiner der eingeschlossenen 33 ProbandInnen gab eine Hörbeeinträchtigung an. In dem finalen erhobenen Datensatz finden sich 21 weibliche Teilnehmerinnen und 12 männliche Teilnehmer. Das diverse Geschlecht wurde von keinem der ProbandInnen gewählt. Alle ProbandInnen gaben Deutsch als Muttersprache an. 32 der

eingeschlossenen ProbandInnen sind Studierende, einE ProbandIn gab eine abgeschlossene Berufsausbildung an. Die Bearbeitungszeit der ProbandInnen betrug zwischen 21 Minuten und 48 Minuten mit einem Mittelwert von 29,52 Minuten. Studierende der Friedrich-Schiller-Universität Jena wurden mit 0,5 Versuchspersonenstunden entlohnt.

### 3.1.2. Stimuli

#### 3.1.2.1. Originale Stimulieraufnahmen

Die verwendeten Stimuli in der Studie waren gemorphte männliche und weibliche Stimmen, eingesprochen in ängstlicher und wütender Emotion. Die Stimuli, die als Ausgangspunkt für das Morphing dienten, stammen aus einer Datenbank von Sascha Frühholz (Department of Cognitive and Affective Neuroscience, Universität Zürich), welche Aufnahmen emotionaler Stimmen, eingesprochen von SchauspielerInnen, enthält. Die enthaltenen Stimuli sind ähnlich zu jenen, die in der Studie von Frühholz et al. (2015) genutzt wurden.

Durch Morphing mit TANDEM STRAIGHT (Kawahara et al., 2008; Kawahara et al., 2013) wurden paarweise Stimuli auf einem Wut-Angst Kontinuum für die gleichen Pseudoworte und SprecherInnen erstellt. Die Stimmen beschreiben auf dem Kontinuum sieben Morphlevel (20-30-40-50-60-70-80) pro SprecherIn und Pseudowort, visualisiert in Abbildung 2.

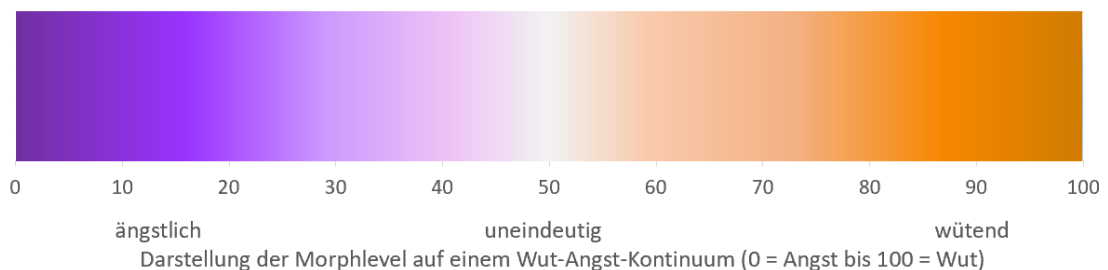


Abbildung 2. Veranschaulichung der Morphlevel auf dem Kontinuum von Angst bis Wut. Die Morphlevel beschreiben verschiedene Grade der jeweiligen Emotionen. Je näher am Mittelpunkt des Kontinuums sich ein Stimulus befindet, desto uneindeutiger ist er. Ist ein Stimulus uneindeutig befindet er sich am PSE (*point of subjective equality*), er kann also keiner Seite des Kontinuums eindeutig zugeordnet werden.

Es werden vier Pseudowörter (/molen/, /namil/, /loman/, /belam/) unterschieden, um eine emotionale Assoziation mit dem semantischen Inhalt der Worte zu vermeiden. Es wurden zwei unterschiedliche Arten von Stimuli erstellt, *Targets* und *Adaptoren*. Die *Adaptoren* befinden sich je an den Enden des Kontinuums, wohingegen die *Targets* sich innerhalb des Kontinuums bewegen.



### 3.1.2.2. *Target*

Die *Targets* sind jene Stimuli, welche von den TeilnehmerInnen der Studie klassifiziert werden sollten. Sie bewegten sich in dem bereits genannten Kontinuum und werden zur Messung der Wahrnehmung der Versuchspersonen verwendet. Es wurden 112 *Targets* erstellt. Insgesamt gab es 4 Sprecheridentitäten (2 männlich, 2 weiblich)  $\times$  7 Morphlevel zwischen Wut und Angst (20-30-40-50-60-70-80)  $\times$  4 Pseudowörter (/molen/, /namil/, /loman/, /belam/), also 112 Targetstimuli.

### 3.1.2.3. *Adaptor*

*Adaptoren* sind jene Stimuli, auf die Versuchspersonen nicht reagieren, sie jedoch aufmerksam anhören sollten. Die *Adaptoren* befinden an den jeweiligen Enden desselben Morphingkontinuums, in unserem Fall zwischen den Emotionen Angst und Wut. Im Gegensatz dazu befinden sich die *Targets* in der Mitte des Kontinuums zwischen den *Adaptoren*. Insgesamt wurden 32 *Adaptoren* erstellt. Hierbei gab es erneut 4 Sprecheridentitäten (2 männlich, 2 weiblich)  $\times$  4 Pseudowörter (/molen/, /namil/, /loman/, /belam/)  $\times$  2 Emotionen (Wut und Angst). Die Emotionen wütend und ängstlich teilten sich in gleichen Teilen auf die *Adaptoren* auf, sodass es je Sprechergeschlecht acht wütende und acht ängstliche Stimuli gab.

### 3.1.3. **Ablauf**

Das Experiment unterteilt sich in einen Baselineblock und zwei Adaptationsblöcke. Im Baselineblock wurden alle 112 Targetstimuli genau einmal präsentiert. Die Aufgabe der ProbandInnen war es, die *Targets* in einer 2-alternative forced choice Aufgabe (2AFC) nach den Emotionen wütend oder ängstlich per Tastendruck zu klassifizieren. Der Baselineblock wurde dabei als Referenz für die späteren Adaptationsbedingungen konzipiert. Zu Anfang sahen die ProbandInnen für 500 ms einen schwarzen Bildschirm, der nur die Zuordnung der Tasten anzeigte. Danach erschien ein grünes Fixationskreuz für 300 ms. Nach diesen 300 ms wurde ein *Target* zufällig abgespielt. Um den Antwortblock zu kennzeichnen, wurde das grüne Fixationskreuz durch ein grünes Fragezeichen ersetzt. Die ProbandInnen hatten 3 Sekunden Zeit, ihre Einschätzung per Tastendruck (f-ängstlich und j-wütend) einzugeben. Nach 3 Sekunden kam die Aufforderung „Bitte reagieren Sie schneller!“, um eine möglichst intuitive Antwort zu erhalten. Zusätzlich zur Antwort wurde die Reaktionszeit der ProbandInnen erfasst. Eine visuelle Darstellung des Baselineblocks befindet sich in Abbildung 3. Um Erschöpfung durch lange Bearbeitungszeiten entgegenzuwirken, hatten die ProbandInnen nach 56 Targetbewertungen die Möglichkeit zu einer Pause,

welche mit einem Tastendruck auf die Leertaste beendet wurde. Der Baselineblock war beendet, wenn alle 112 *Targets* genau einmal bewertet wurden. Im Vorfeld hatten die ProbandInnen die Möglichkeit den Ablauf in einem Übungsblock zu testen.

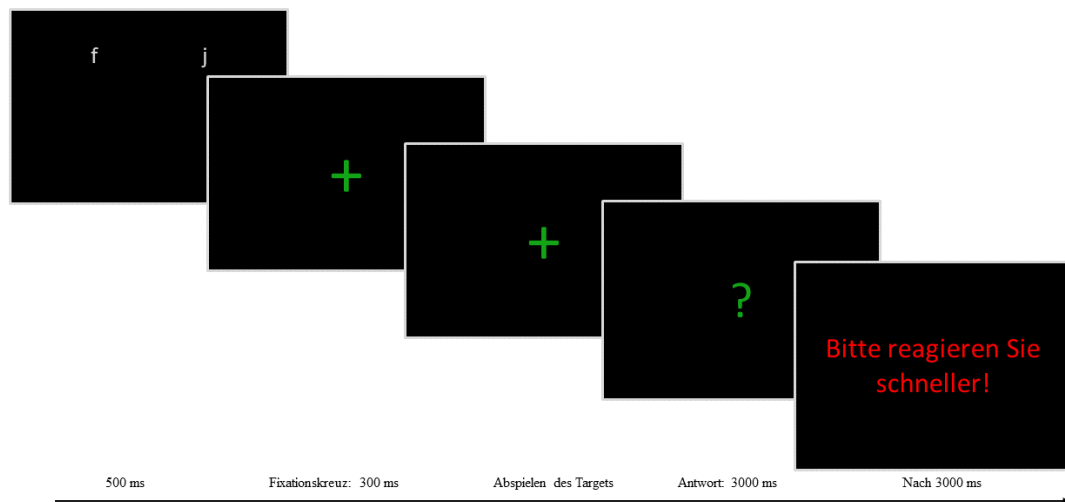


Abbildung 3. Schematische Visualisierung des Ablaufs der Baselineblocks.

Die Adaptation selbst wurde in zwei verschiedenen Adaptationsbedingungen mit zwei unterschiedlichen Adaptationstypen durchgeführt. In einer Bedingung wurde auf wütende weibliche und ängstliche männliche Stimmen adaptiert. In der anderen Bedingung wurde entgegengesetzt auf ängstliche weibliche und wütende männliche Stimmen adaptiert. Die Adaptation fand dabei immer gleichzeitig statt durch eine durchmischte Präsentation männlicher und weiblicher Stimmen. Wichtig ist, dass in einer Adaptation zeitgleich auf wütende und ängstliche Stimuli adaptiert und dabei das Sprechergeschlecht zusätzlich dazu präsentiert wurde. Das heißt, es fand eine Adaptation auf die Emotionen Wut und Angst statt, aber je auf ein spezifisches Sprechergeschlecht. Im Adaptationsblock wurde beginnend ein rotes Fixationskreuz ohne die Zuordnung der Tasten für 300 ms präsentiert. Danach wurden die gesamten 16 *Adaptoren* (2 Sprecheridentitäten  $\times$  4 Pseudowörter  $\times$  2 Emotionen (Wut/Angst)) je Adaptation präsentiert. Zwischen jedem *Adaptor* wurde eine Pause von 100 ms gelassen. Nach der Adaptation erschienen für 3 Sekunden die Worte „Jetzt geht es los!“ in grüner Schriftfarbe, um für die ProbandInnen das Ende der Adaptation und den Beginn der erneuten Targetbewertung zu kennzeichnen. Der Aufbau der Targetbewertung entspricht dem des Baselineblocks. Nach 4 Targetbewertungen im Adaptationsblock wurden zwei *TopUp-Adaptoren* abgespielt, um über die Zeit der Bewertung die Adaptation aufrecht zu erhalten. Dazu wurde für 300 ms ein rotes Fixationskreuz und danach zwei der im Adaptationsteil abgespielten *Adaptoren* erneut präsentiert, jedes Sprechergeschlecht einmal. Die *TopUp-Adaptoren* wurden immer

zufällig aus den *Adaptorstimuli* ausgewählt, wobei die Reihenfolge ebenfalls randomisiert wurde. Nach den *TopUp-Adaptoren* ging es wieder in die 2AFC derselben *Targets* über, wie im Baselineblock. Dies wurde wiederholt, bis alle 112 *Targets* erneut bewertet wurden. Zum bessern Verständnis befindet sich eine schematische Darstellung des Adaptationsblocks in Abbildung 4. Genau wie im Baselineblock hatten die ProbandInnen nach 56 Targetbewertungen die Möglichkeit für eine Pause. Vor der Durchführung der Adaptation wurde der Ablauf mit einem Übungsblock veranschaulicht.

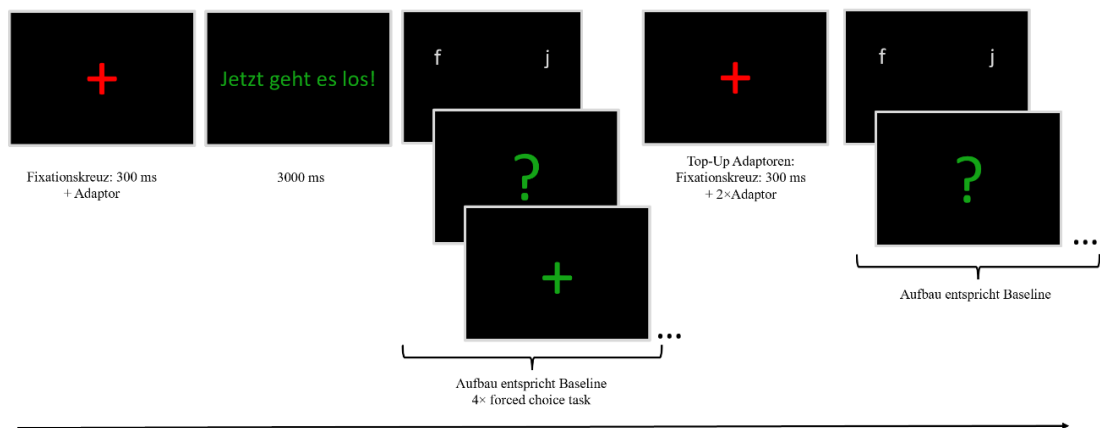


Abbildung 4. Schematische Visualisierung des Ablaufs des Adaptationsblocks.

Die Studie wurde als within-Design durchgeführt, daher nahm jede ProbandIn an beiden Adaptationsblöcken teil. Die Reihenfolge der Adaptationsblöcke wurde bei den ProbandInnen zufällig konzipiert, um eine Rückführung des Effekts auf die Reihenfolge der Adaptationen auszuschließen. Alle ProbandInnen führten zuerst den Baselineblock durch, danach folgte zufällig eine der Adaptationen und im Anschluss die erneute Bewertung der *Targets*, wie im Baselineblock, wobei zusätzlich die zwei *TopUp-Adaptoren* nach 4 Targettrails abgespielt wurden. Nach diesem ersten Adaptationsblock wurde der zweite Adaptationsblock auf die gleiche Art und Weise durchgeführt, nur mit der entgegengesetzten Adaptation, welche im ersten Adaptationsblock nicht durchgeführt wurde.

Vor und nach der Durchführung der Studie erhielten die ProbandInnen einige qualitative Fragen. Die Fragen vor der Studie bezogen sich auf Alter, Studiengang, Muttersprache und Höreinschränkungen. Nach Abschluss des Experiments erhielten die ProbandInnen Fragen zur Verständlichkeit der Studie, Entwicklung von Antwortstrategien und erhielten Möglichkeit zum Stellen von Fragen oder Anregungen.

### 3.1.4. Erhebung

Genau wie das Experiment von Bestelmeyer et al. (2010a) wurde die Studie in einem Onlineformat durchgeführt (Stoet, 2010, 2017). Dies wurde mit Hilfe von PsyToolkit realisiert (Stoet, 2010, 2017). Der Erhebungszeitraum betrug 4 Wochen.

### 3.2. Auswertung

Zur Analyse der Daten wurde eine Auswertung mit dem Statistikprogramm R durchgeführt. Baselineblock und Adaptationsblöcke wurden getrennt voneinander ausgewertet. Die Auswertung beider Blöcke wurde mit einem Datensatz von  $N = 33$  erstellt. Es wurden 7 ProbandInnen der ursprünglichen 40 entfernt. Die ausgeschlossenen ProbandInnen wiesen entweder im Baselinedatensatz oder im Adaptationsdatensatz ein  $R^2 < .70$  und einen  $20 > PSE > 80$  vor. Der PSE kann sich ausgehend von seiner Definition nur im Bereich der Morphlevel bewegen, daher wurden die ProbandInnen, die dem nicht entsprachen von der Analyse ausgeschlossen. Es wurde je ein Modell der kumulativen Gaussfunktion für die Adaptationsblöcke ( $2 \times$  Sprecher-geschlecht  $\times 2$  Adaptationstyp) und den Baselineblock ( $2 \times$  Sprecher-geschlecht), gemittelt und für jedeN ProbandIn einzeln berechnet.

#### 3.2.1. Ergebnisse der Adaptationsblöcke

##### 3.2.1.1. Adaptationsblock: Analyse des point of subjective equality (PSE)

Zur statistischen Auswertung wurde zur Analyse des „Anteils der ‚wütend‘-Antworten“ eine  $2 \times 2$  within-subject-ANOVA durchgeführt mit dem PSE als abhängige Variable und dem Adaptationstyp (w\_ängstlich/m\_wütend und w\_wütend/m\_ängstlich), sowie dem Sprecher-geschlecht (männlich und weiblich) als within-Faktoren. Dabei ergab sich ein signifikanter Interaktionseffekt zwischen Adaptationstyp und Sprecher-geschlecht ( $F(1, 32) = 108.05, p < .001, \eta_p^2 = .772$ ). Die beiden Haupteffekte nach Sprecher-geschlecht ( $F(1, 32) = 3.07, p = .089, \eta_p^2 = .088$ ) und Adaptationstyp ( $F(1, 32) = 0.02, p = .888, \eta_p^2 < .011$ ), also w\_ängstlich/m\_wütend und w\_wütend/m\_ängstlich, sind nicht signifikant.

Der signifikante Interaktionseffekt der ANOVA wurden in einem post-hoc-Test nachgetestet. Dazu wurden gepaarte t-Test berechnet. Der Adaptationstyp wurde hierbei konstant gehalten und der Effekt des Sprecher-geschlechts mit gepaarten t-Tests nachgetestet. Dabei zeigte sich für beide Adaptationstypen ein signifikanter Effekt, jedoch in entgegengesetzter Richtung: Während bei der „w\_ängstlich/m\_wütend“-Adaptation der PSE bei weiblichen *Targets* kleiner war als bei männlichen ( $t(32) = -3.30, p < .001$ ), war das bei der „w\_wütend/m\_ängstlich“-Adaptation andersherum

( $t(32) = 5.91, p < .001$ ). Dies spricht dafür, dass in den Blöcken die Emotion geschlechterspezifisch adaptiert wurde (siehe Abbildung 5).

Auf Grundlage der signifikanten Interaktion kann geschlossen werden, dass sich der PSE durch die verschiedenen Adaptationstypen unterschiedlich verschiebt, je nach Sprecher-geschlecht, was für die Belegung der gestellten Hypothese spricht.

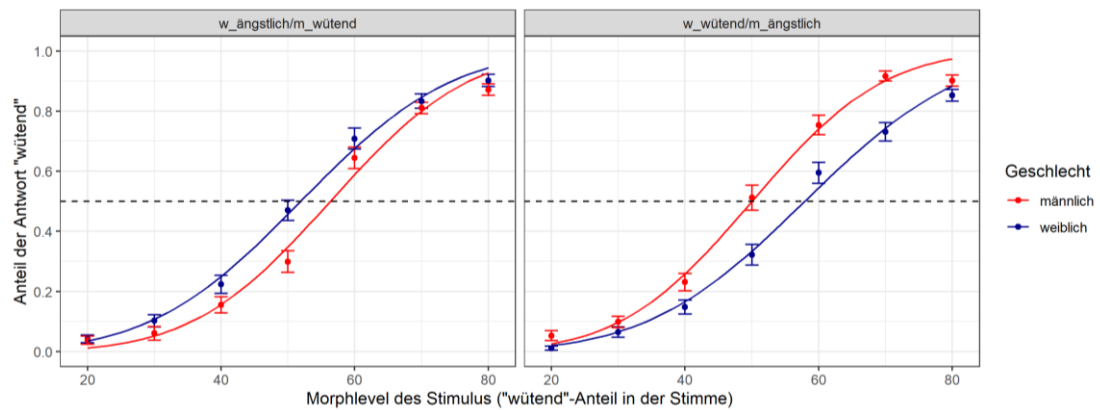


Abbildung 5. Darstellung der Bewertung der Targetstimuli aus den Adaptationsblöcken als kumulative Gaussfunktion getrennt nach Adaptationstypen und Sprecher-geschlecht. (gemittelte Daten; N=33).

In dieser Darstellung lässt sich die geschlechterspezifische Verschiebung des PSE abhängig von der Adaptation erkennen, welche durch die signifikante Interaktion aus der ANOVA nahegelegt wird. Der Effekt der Adaptation lässt sich an der Verschiebung des PSE erkennen. Der PSE ist in dieser Grafik an der Schnittstelle aus den farbigen Graphen und der gestrichelten Linie abzulesen. Die roten Graphen beschreiben die Bewertung der männlichen Stimuli. Im Adaptationstyp „w\_ängstlich/m\_wütend“ verschiebt sich der PSE der männlichen Stimuli in Richtung „wütend“, das heißt, es benötigt nach der Adaptation einen höheren „wütend“-Anteil in der Stimme, um von den ProbandInnen als wütend wahrgenommen zu werden. Im Adaptationstyp „w\_wütend/m\_ängstlich“ verschiebt sich der PSE der männlichen Stimuli in Richtung „ängstlich“, was bedeutet, dass der „wütend“-Anteil geringer ist, wenn der Stimulus bereits als wütend klassifiziert wird. Dasselbe Muster zeichnet sich ebenfalls bei der Bewertung der weiblichen Stimuli, welche durch den blauen Graphen beschrieben sind, ab. Im Adaptationstyp „w\_ängstlich/m\_wütend“ wurden die weiblichen Stimuli auf Angst adaptiert, sodass die ProbandInnen mit einem geringeren „wütend“-Anteil in der Stimme einen Stimulus als wütend einschätzten. Dieser Effekt zeigt sich im Adaptationstyp „w\_wütend/m\_ängstlich“ entgegengesetzt, womit ein größerer „wütend“-Anteil in der Stimme nötig ist, dass die Emotion als Wut erkannt wird.

### 3.2.1.2. Adaptationsblock: Analyse der Standardabweichung

Um auszuschließen, dass die Steigung der Kurve von der durchgeführten Manipulation abhängt, wurde eine  $2 \times 2$  within-subject-ANOVA durchgeführt mit der Standardabweichung des Antwortverhaltens der ProbandInnen als abhängige Variable und dem Adaptationstyp (w\_ängstlich/m\_wütend und w\_wütend/m\_ängstlich) und dem Sprechergeschlecht (männlich und weiblich) als within-Faktoren. Die Haupteffekte auf Sprechergeschlecht ( $F(1, 32) = 2.9, p = .098, n_p^2 = .083$ ) und den Adaptationstyp ( $F(1, 32) = 0.19, p = .668, n_p^2 < .016$ ) sind nicht signifikant. Im Haupteffekt auf das Sprechergeschlecht zeigt sich ein leichter Trend, der an dieser Stelle allerdings nicht groß genug ist, um aussagekräftig zu sein. Die Interaktion zwischen den within-Faktoren Sprechergeschlecht und Adaptationstyp ist ebenfalls nicht signifikant ( $F(1, 32) = 1.955, p = .172, n_p^2 = .058$ ). Dies validiert, dass die Steigung der kumulativen Gaussfunktion durch das Antwortverhalten der ProbandInnen und nicht durch die Manipulation verursacht wird.

### 3.2.2. Ergebnisse des Baselineblocks

#### 3.2.2.1. Baseline: Analyse des point of subjective equality (PSE)

Zur Analyse des Baselineblocks wurden gepaarte t-Tests durchgeführt. Bei der Erstellung des Plots mit der Funktion ggplot ist aufgefallen, dass auch schon im Baselineblock die Sprechergeschlechter unterschiedlich wahrgenommen werden.

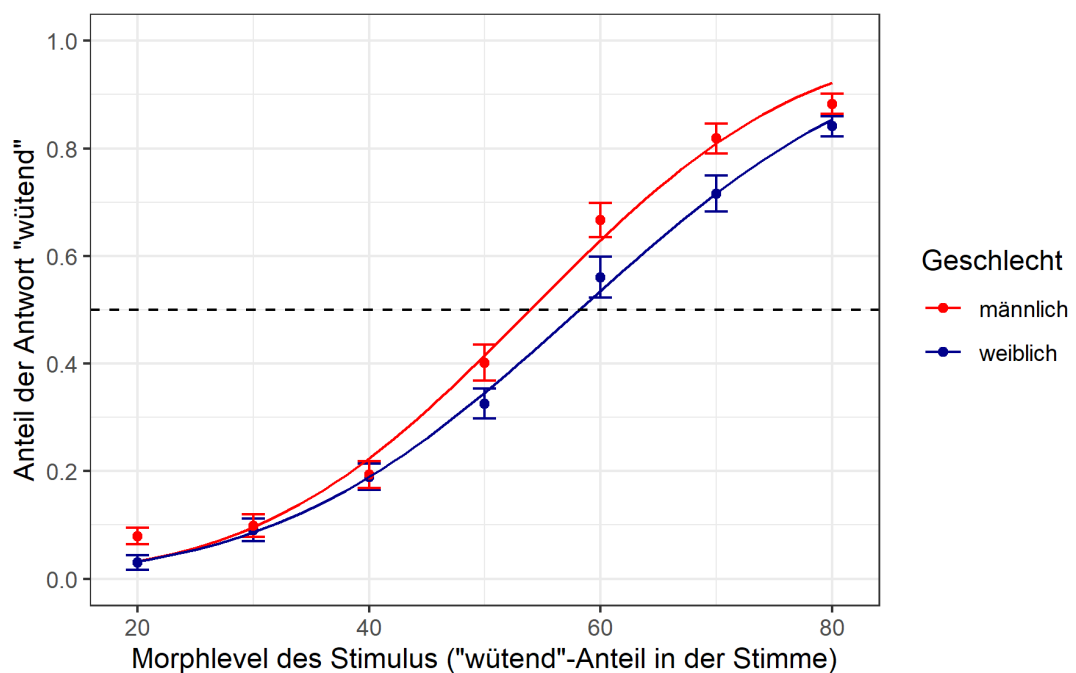


Abbildung 6. Darstellung der Bewertung der Targetstimuli aus dem Baselineblock als kumulative Gaussfunktion. Die Männer wurden als wütender wahrgenommen als die Frauen. (gemittelte Daten;  $N=33$ ).

Mit einem t-Test sollte eruiert werden, ob der Unterschied des PSE zwischen den Sprecher\*geschlechtern signifikant ist. Der gepaarte t-Test ergab eine signifikante Differenz des PSE zwischen den Gruppen „männlich“ und „weiblich“ ( $t(32) = 2.48$ ,  $p=.019$ ).

In Abbildung 6 lässt sich erkennen, dass in der Bewertung der *Targets* im Baselineblock die Männer (roter Graph) als vermehrt wütend und die Frauen (blauer Graph) eher als ängstlich wahrgenommen wurden. Das lässt sich daran erkennen, dass die Baseline bei den Männern früher die gestrichelte Linie kreuzt, was einen unterschiedlichen PSE für die Geschlechter anzeigt. Männliche und weibliche Stimuli werden folglich auch ohne Adaptation signifikant unterschiedlich bewertet.

#### 3.2.2.2. *Baseline: Analyse der Standardabweichung*

Es wurde zusätzlich ein t-Test zwischen der Standardabweichung des Antwortverhaltens der ProbandInnen und dem Sprecher\*geschlecht durchgeführt. Hierbei ergab sich kein signifikanter Unterschied in der Bewertung zwischen Männern und Frauen ( $t(32) = 1.29$ ,  $p=.208$ ).

#### **4. Diskussion**

Das Ziel dieser Studie war es, die Ergebnisse von Bestelmeyer et al. (2010a) für emotionale Gesichtsausdrücke in emotionalen Stimmen zu replizieren und die Fragestellung zu beantworten, ob bei der Wahrnehmung von Emotionen in Stimmen geschlechterspezifische Adaptation stattfindet. Die gefundenen Ergebnismuster sprechen für eine geschlechterspezifische Adaptation bei Emotionen in der Stimme, da für die PSEs eine signifikante und hypothesenkonforme Interaktion zwischen Adaptationstyp und Sprechergeschlecht beobachtbar waren. Wurden männliche Gesichter auf „wütend“ und weibliche Gesichter auf „ängstlich“ adaptiert, so wurden männliche Targetgesichter eher als „ängstlich“ und weibliche Gesichter als „wütend“ klassifiziert. Ein umgekehrtes Muster fand sich bei umgedrehter Zuordnung von Geschlecht zu Emotion im Adaptationsblock. Hätten die ProbandInnen nicht geschlechterspezifisch adaptiert, hätten sich die Adaptationseffekte gegenseitig aufheben müssen, da in den Adaptationsblöcken gleich viele wütende und ängstliche Stimuli gezeigt wurden. Da dies nicht der Fall ist, spricht das für eine geschlechterspezifische Adaptation auf die jeweilige Emotion. Die Ergebnisse decken sich somit mit denen des Experiments von Bestelmeyer et al. (2010a) zum Einfluss des Geschlechts auf die Wahrnehmungseffekte von emotionalen Gesichtsausdrücken. Wurde gleichzeitig adaptiert und ein Geschlecht auf wütend und das andere auf ängstlich, so wurde das erste Geschlecht weniger als wütend wahrgenommen und das andere vermehrt als wütend (Bestelmeyer et al., 2010a). Die Daten aus dieser Studie wurden in die Modalität der Stimme übertragen, sodass die Fragestellung mit dem durchgeführten Experiment beantwortet werden konnte. Das bedeutet, die Adaptation erzielte den erwünschten Effekt.

Die hier durchgeführte Studie ist nicht die erste, welche sich mit den geschlechterspezifischen Effekten auf die Adaptation von Emotionen beschäftigt. Die geschlechterspezifische Adaptation konnte bereits in Gesichtszügen (Webster et al., 2004), aber auch bei emotionalen Gesichtern (Bestelmeyer et al., 2010a) nachgewiesen werden. Nussbaum et al. (2022) fanden ebenfalls Hinweise darauf, dass die Adaptation auf emotionale Stimmen vom Sprechergeschlecht moduliert wird.

In der zwischenmenschlichen Kommunikation ist es wichtig, auch schon kleine Unterschiede in der emotionalen Verfassung des Gegenübers wahrzunehmen, um angemessen auf eine mögliche Veränderung der Stimmung zu reagieren. Dafür ist eine differenzierte Wahrnehmung unterschiedlicher Personen und ihrer emotionalen



Zustände nötig. Diese Veränderung in der emotionalen Lage des Gegenübers - beispielsweise, wenn ein Freund während eines lustigen gemeinsamen Zusammensitzens eine schlechte Nachricht bekommt - wahrzunehmen, ist bedeutend für den korrekten Umgang mit diesen Situationen. Wesentlich ist dabei zu beachten, dass es vordergründig wichtiger ist, die Veränderungen in den Emotionen zu erkennen, als die absolute Emotion zuzuordnen. Das Gehirn betrachtet demnach eher die relativen Variationen der Emotionen, wenn es darum geht, die Veränderung einer Emotion zu erfassen. In einer neueren Studie von Nussbaum et al. (2022) wird dieses Erkennen einer relativen Veränderung in der emotionalen Atmosphäre mit der Adaptation insofern in Verbindung gesetzt, dass es durch die Adaptation zu einer Sensibilisierung für emotionale Veränderungen kommt. Kleinere Unterschiede in der allgemein herrschenden Emotion, sei es Freude, Angst oder Ekel, werden demnach schneller wahrgenommen, da die emotionale Atmosphäre einen Einfluss auf die Wahrnehmung der Emotionen hat (Bestelmeyer et al., 2010b; Nussbaum et al., 2022).

Damit nun diese kleineren Veränderungen wahrgenommen werden können, ist es naheliegend, dass möglichst viele Informationen über das Gesprächsgegenüber wahrgenommen und neuronal verarbeitet werden, darunter auch das Geschlecht. Dies zeigt, dass es eine differenzierte Verarbeitung verschiedener Informationen geben muss, welche sich durch das durchgeführte Adaptationsparadigma veranschaulichen lässt. Infolgedessen kann die Wahrnehmung neu aufgenommener Reize durch die bereits aufgenommene Information beeinflusst werden. Im Falle dieser Studie führt dies zu einer veränderten Targetbewertung. Die Studie lässt somit Spekulationen auf die neuronale Verarbeitung von Emotionen zu. Es liegt aufgrund des Ergebnisses einer geschlechterspezifischen Adaptation nahe, dass Emotion und Geschlecht, wenn beide Informationen über die Stimme aufgenommen werden, nicht einfach nur verarbeitet werden, sondern miteinander interagieren. Dieses Ergebnis lässt sich in die bisher erforschten Zusammenhänge der Verarbeitung von Geschlecht und Emotionen einordnen. Das Geschlecht zählt zu den extralinguistischen Signalen, das heißt es ist auch in einer Person relativ stabil (Schweinberger et al., 2014). Das wiederum bedeutet, dass sich diese Informationen innerhalb einer zwischenmenschlichen Interaktion nicht ändern (Schweinberger et al., 2014), anders als die Emotionen, welche zu den paralinguistischen Signalen gehören, also jene Signale, die sich situationsspezifisch unterscheiden können und sich auch in einer Interaktionsphase verändern können (Schweinberger et al., 2014). Das biologische Geschlecht verändert

sich beim Menschen nicht situationsspezifisch. Somit ist die Information, welche eine Person über die Stimme preisgibt, gleichbleibend. In Folge dessen entsteht die Theorie, dass Emotionen und Geschlecht zunächst in verschiedenen Pfaden verarbeitet werden, da es sich unterschiedlich verändernde Signale sind, welche jedoch miteinander interagieren können (Belin et al., 2011). Diese Theorie steht im Einklang mit den Ergebnissen der hier durchgeführten Studie, da das Geschlecht im Baselineblock auch bereits ohne die Adaptation einen Einfluss auf die Wahrnehmung der Emotionen hat. Gäbe es keinen Informationsaustausch auf den beiden Verarbeitungsbahnen, so dürfte die Adaptation eigentlich keinen Einfluss auf die Wahrnehmung der Emotionen haben. Die Datenlage in der hier durchgeführten Studie ist somit im Prinzip deckungsgleich mit den Ergebnissen aus Nussbaum et al. (2022), da die Effekte nach der Veränderung des Geschlechts zwischen *Target* und *Adaptoren* beinahe nicht mehr vorhanden waren, was ebenfalls nahelegt, dass Emotion und Geschlecht nicht unabhängig voneinander verarbeitet werden.

Die Erforschung der neuronalen Abläufe der Verarbeitung von Emotionen ist durch die Tatsache erschwert, dass eine alleinige Betrachtung der Emotionsverarbeitung unabhängig von anderen Eigenschaften der Stimme nur schwer möglich ist (Young et al., 2020). Diese Komplexität wird von den Daten der Studie insofern unterstützt, dass eine Interaktion in der Verarbeitung aus Emotionen und Geschlecht aus der Stimme zu beobachten ist, was in Einklang steht mit der Annahme einer multimodalen neuronalen Verarbeitung, dem Modell, das sich in der neurologischen Forschung andeutet (Brück et al., 2011). Was man mit dieser Studie demnach aussagen kann: Emotionen in Stimmen werden in Abhängigkeit des Geschlechts verarbeitet.

Mit den Ergebnissen der Studie wird davon ausgegangen, dass die Unterschiede der Wahrnehmung der Emotionen auf die Gruppen der Geschlechter „männlich“ und „weiblich“ zurückgeführt werden können. Betrachtet man alternative Erklärungen für diesen Effekt, so könnte man argumentieren, dass die Gruppen „weiblich“ und „männlich“ keine Geschlechterspezifität darstellen, sondern die Unterscheidung durch einen anderen Parameter stattfindet.

Es könnte angenommen werden, dass es anstatt einer geschlechterspezifischen Adaptation zu identitätsspezifischen Adaptationseffekten gekommen ist. Dies könnte der Fall sein, wenn genau zwei Personen, eine männlich, die andere weiblich für das Erstellen der *Targets* und *Adaptoren* genutzt wurden. So könnte man argumentieren, dass hier die Identität der Personen ausschlaggebend war für die unterschiedlichen

Adaptationseffekte bei Frauen und Männern, da die Adaptation auf Identität bereits nachgewiesen werden konnte (Schweinberger et al., 2008). Für die Stimuli wurden in unserer Studie allerdings je Geschlecht 2 Sprecheridentitäten genutzt und nicht nur eine Identität. Somit kann der Einfluss einer einzelnen Identität auf die Adaptation umgangen und die Effekte können eher auf die Geschlechter zurückgeführt werden. Eine weitere alternative Erklärung für die Gruppenzusammenstellung ist über die Parameter in der Stimme. Es wurden die verschiedenen Parameter thematisiert, welche Informationen in der Stimme transportieren, unter anderem das Geschlecht, was sich, auf Grundlage der Theorie von Banse und Scherer (1996), über akustische Parameter in der Stimme äußert und kodiert wird (Banse & Scherer, 1996; Juslin & Laukka, 2003; Scherer, 1986). Es wurde angenommen, dass die Grundfrequenz F0, aufgrund der deutlichen Unterschiede bei Männern und Frauen, der Hauptparameter ist, um Geschlechter in der Stimme zu unterscheiden (Schweinberger et al., 2008; Skuk & Schweinberger, 2013). Ausgehend davon könnte die Aussage getroffen werden, dass die Unterschiede zwischen den Gruppen auf den verschiedenen Grundfrequenzen F0 resultieren (Männer: 100 bis 150 Hz, Frauen: 190 bis 250 Hz (Kiese-Himmel, 2016)). Inwieweit diese akustischen Unterschiede zwischen männlichen und weiblichen Stimmen für die geschlechterspezifischen Effekte verantwortlich sind, wird mit dieser Studie nicht erforscht. Eine mögliche Erklärung für die geschlechterspezifischen Effekte sein könnte sein, dass aufgrund der großen Unterschiede der Grundfrequenz zwischen Männern und Frauen die Verarbeitung der Informationen in zwei verschiedenen „voice spaces“ stattfindet (Lima & Castro, 2011). Alle Stimmen mit einer geringeren F0 werden auf der einen Seite des mentalen Raums verarbeitet und Stimmen mit einer größeren F0 auf der anderen Seite des mentalen Raums (Lima & Castro, 2011). Damit hätten männliche und weibliche Stimmen eine durch die akustischen Parameter unterschiedene mentale Repräsentation, was für eine getrennt ablaufende Adaptation in den jeweiligen „voice spaces“ (Lima & Castro, 2011) der Stimuli spricht, wodurch dann eine geschlechterspezifische Adaptation zustande kommen könnte. Anknüpfend wäre interessant zu sehen, welchen Einfluss der Kontext auf eine Wahrnehmung der Informationen aus der Stimme hat, da der Kontext die Wahrnehmung beeinflussen kann (Frühholz & Schweinberger, 2021). Kann der Kontext Einfluss nehmen auf die „voice spaces“ (Lima & Castro, 2011)? Kann der Kontext zu einer weniger klaren Unterscheidung der Geschlechter führen? Wann wird eine Stimme als männlich und wann eher als wütend wahrgenommen?

Interessant ist zusätzlich, dass es bereits ohne eine Adaptation einen Unterschied in der Wahrnehmung der Geschlechter gibt. Schon im Baselineblock werden Männer und Frauen in ihrer Emotionalität unterschiedlich bewertet, wobei Frauen als ängstlicher und Männer als wütender wahrgenommen wurden, auch wenn zu diesem Zeitpunkt noch keine Adaptation stattgefunden hatte, welche die Wahrnehmung beeinflussen könnte. Die gleichen Ergebnismuster der Baseline finden sich in Nussbaum et al. (2022). Umso beeindruckender ist es, dass es in dem Adaptationstyp, in welchem die männlichen Stimuli auf ängstlich und die weiblichen Stimuli auf wütend adaptiert wurden, möglich war, einen robusten signifikanten kontrastiven Nacheffekt der Adaptation zu finden, da hier gegen die ursprüngliche Wahrnehmung der Emotionen gearbeitet werden musste. Die Tendenz, welche sich im Baselineblock in der Wahrnehmung der Emotionen ohne Adaptation zeigt, dass Frauen eher als ängstlich und Männer eher als wütend wahrgenommen werden, musste überschrieben werden, um diesen Effekt in der entgegengesetzten geschlechterspezifischen Adaptationsbedingung umzudrehen.

Nussbaum et al. (2022) stellt für das Phänomen der unterschiedlichen Bewertungen der Geschlechter in der Baseline zwei mögliche Erklärungsansätze auf, wobei eine Erklärung sich auf die Seite des Hörers und die andere sich auf die Seite des Sprechers bezieht. Betrachtet man die Seite des Hörers in Bezug auf die unterschiedliche Wahrnehmung der Emotionen bei Männern und Frauen auch ohne die Adaptation, so lässt sich dies möglicherweise auf das stereotypische Wissen der Hörer über die Geschlechter zurückführen (Nussbaum et al., 2022). Männer werden in ihrer sozialen Rolle eher mit „Dominanz“ oder „Aggression“ verbunden, wohingegen Frauen eher die Rollen mit den Thematiken „Unterwürfigkeit“ oder „Ängstlichkeit“ zugeordnet werden (Nussbaum et al., 2022). Wird nun einer Person die Aufgabe gegeben, Stimuli in die Kategorien Wut und Angst einzuordnen, so ist es naheliegend, dass die Geschlechter nach stereotypischen Mustern klassifiziert werden. Wut wird im Alltag eher mit einer lauten Stimme assoziiert, welche charakteristisch eher bei Männern zu finden ist, Frauen werden häufiger mit einer leiseren Stimme verbunden, welche sich stereotypisch in die zurückhaltende, ängstliche Richtung einfügt. Der zweite Erklärungsansatz bezieht sich auf die Sprecherseite der Kommunikation. Hiernach könnten Männer besser in der Lage sein, Wut auszudrücken, wohingegen Frauen besser in der Lage sind, Angst über die Stimme zu transportieren (Nussbaum et al., 2022).

Im sozialen Kontext lernen wir Geschlechter zu unterscheiden. Diese Fähigkeit wird besonders ab der Pubertät relevant und prägt sich weiter aus, da in diesem Lebensabschnitt die akustischen Unterschiede zwischen den Geschlechtern männlich und weiblich deutlicher werden (Schweinberger et al., 2014). Brody (2000) nimmt dies in seinem Modell als Ausgangspunkt, wonach biologisch basierte Vorannahmen und Eigenschaften im Prozess der Sozialisation eher Frauen oder eben eher Männern zugeordnet werden (Brody, 2000). Aus diesem Modell folgt, dass Frauen als emotional ausdrucksstärker wahrgenommen werden, wobei sie zusätzlich Positives vermehrt nach außen zeigen und negative Emotionen internalisieren (Brody, 2000; Nussbaum et al., 2022). Männer hingegen zeigen insgesamt weniger Emotionen, neigen allerdings zugleich dazu, eher negative Emotionen zu zeigen, wie zum Beispiel Wut (Nussbaum et al., 2022). Dies legt nahe, dass die Geschlechter Emotionen auf unterschiedliche Art und Weise ausdrücken, sodass die Unterschiede in der Wahrnehmung der Emotionen bereits ohne Adaptation zwischen den Geschlechtern logisch erscheinen.

Betrachtet man dies nun im Zusammenhang mit den in dieser Studie gefundenen Ergebnissen, so fällt auf, dass den Frauen vermehrt die negative Emotion „Angst“ zugeordnet wurde. Dies steht nicht im Einklang mit der Theorie nach Brody (2000), wonach Frauen solche Emotionen eher nicht nach außen zeigen. Dies könnte durch die Auswahl an Emotionen zustande kommen. Die ProbandInnen hatten nur die Auswahl zwischen „Wut“ und „Angst“ als Emotionen, Männer zeigen im sozialen Kontext eher Wut, werden daher mit Wut assoziiert, was dafür sorgt, dass die Kombination von „weiblich“ und „Angst“ verbleibt. Es wäre interessant zu untersuchen, was passieren würde, wenn andere Emotionen genutzt werden, wie beispielsweise ein Wut-Freude-Kontinuum. Finden sich ähnliche große Unterschiede in der Wahrnehmung ohne eine Adaptation der Wahrnehmung oder verändern sich die Effekte?

Jeder dieser Ansätze bringt valide Punkte, welche die Unterschiede in der Wahrnehmung der Emotionen zwischen den Geschlechtern erklären könnten.

#### **4.1. Limitationen der Studie**

Die Studie zeigt einige Limitationen auf. Die ProbandInnen sind vor allem akademische TeilnehmerInnen, was bei Studien dieser Art häufig der Fall ist. Der Online-Link zur Studie wurde über Mailinglisten in studentischen Kreisen verbreitet, wodurch es eben zu einer Vielzahl an akademischen TeilnehmerInnen kommt.

Zudem wurde die Studie als Online-Studie durchgeführt. Diese Entscheidung wurde bewusst so getroffen, um die Ergebnisse mit denen von Bestelmeyer et al. (2010a)

vergleichbar zu machen, welche das Experiment der Studie ebenfalls als Online-Studie durchgeführt hatten. Dies hat allerdings den Nachteil, dass mögliche Störfaktoren, wie beispielsweise Unterbrechungen durch weitere Haushaltspersonen, ablenkende Gerüche oder ähnliches, nicht kontrolliert werden können, sowie die Unterschiede in der technischen Ausstattung der ProbandInnen, wie zum Beispiel Computer oder Kopfhörer. Um dem entgegenzuwirken, wurden die ProbandInnen aufgefordert, sich einen ruhigen Ort zu suchen und die Studie mit Kopfhörern durchzuführen. Dies kann bei den genannten Problemen helfen, ist allerdings kein Vergleich zu kontrollierten Laborbedingungen. In Online-Studien ist nicht auszuschließen, dass die Instruktionen zum Teil fehlerhaft befolgt oder ignoriert werden.

Es gab es einen Dropout von 16 ProbandInnen. Der Link wurde 56-mal aufgerufen, allerdings wurde die Studie von nur 40 ProbandInnen beendet. Dieser Dropout ist über mögliche technische Schwierigkeiten, wie Internetprobleme, zu erklären. Auch ein eigenständiger Abbruch der Studie kann Ursache für die Dropouts sein, da den ProbandInnen vor Beginn der Studie mitgeteilt wurde, dass sie die Teilnahme jederzeit ohne Angabe von Gründen abbrechen können. Trotz alledem konnten mit der Studie robuste Ergebnisse erzielt werden, sodass ein Online-Format für diese Studie geeignet gewesen ist.

Ein weiterer Kritikpunkt ist, dass es bei der Länge der Studie zu Erschöpfung kommen kann. Im Ablauf der Studie waren immer wieder Pausen vorgesehen, die die ProbandInnen selbstständig nach Bedarf beenden konnten. Die selbstständige Pauseneinteilung könnte allerdings die Chance für mögliche Ablenkungen erhöhen.

Eine Stärke der Studie ist der übernommene Aufbau von Bestelmeyer et al. (2010a). Mit diesem komplexen Aufbau der zeitgleichen Adaptation können Einflüsse unterschiedlicher Parameter auf die Adaptation gemessen werden, womit Rückschlüsse auf eine gemeinsame oder getrennte neuronale Verarbeitung möglich werden.

#### **4.2. Weiterführende Forschung und folgende Fragestellungen**

Die anknüpfende Forschung an diese Ergebnisse könnte analysieren, wie das Encoding der Emotionen und anderen Informationen, wie beispielsweise das Geschlecht, in der Stimme funktioniert. Was genau sorgt dafür, dass wir jemanden als Mann oder Frau erkennen, und wie beeinflusst das anschließend die Emotionswahrnehmung? Weshalb gerade diese Forschung interessant ist, zeigt sich in den Daten aus dem Baselineblock. Schon hier, ohne Adaptation, werden Männer und Frauen in den Emotionen

unterschiedlich bewertet und wahrgenommen. Männer werden eher als wütender wahrgenommen und Frauen als ängstlicher. Es wäre spannend zu untersuchen, woher dieses Muster kommt und welche Parameter dafür verantwortlich sind. Fant (1971) liefert mit der Source-Filter-Theorie eine gute Grundlage für diese Forschung, auf der man die Zusammensetzung der verschiedenen Informationen, welche durch die Stimme transportiert werden können, erklären kann. Durch diese Theorie liegt es nahe, dass unterschiedliche Muskelspannungen und Stellungen der Artikulatoren die stimmlichen Informationen beeinflussen können, es gibt aber auch Hinweise darauf, dass andere Teile der Stimme, welche beispielsweise durch das Lungenvolumen bestimmt werden, relativ stabil bleiben. Die Nuancen unterschiedlicher Zusammenhänge bleiben weiter zu erforschen. Banse und Scherer (1996) haben in einem ersten Ansatz versucht, die Muster der Informationsvermittlung über die Stimme aufzuschlüsseln. Dabei wurden Annahmen gemacht, wie eine mögliche Zusammensetzung der akustischen Signale aussehen könnte, was Banse und Scherer (1996) als *stimulus evaluation checks* (SECs) betiteln. Die SECs bilden typische Zusammensetzungen für bestimmte Emotionen (Banse & Scherer, 1996). In aktueller Forschung wird versucht, diese Annahmen der akustischen Zusammensetzung der Emotionen zu überprüfen. Für die Klassifikation von Emotionen finden sich Hinweise darauf, dass die Grundfrequenz eine größere Rolle spielt (Juslin & Laukka, 2003). Neuere Untersuchungen beschäftigen sich damit, welche Parameter in der Stimme konkret die Information der Emotionen transportieren. Im Vergleich von Grundfrequenz und Klangfarbe konnte ermittelt werden, dass die Klangfarbe einen größeren Anteil für die Effekte der Adaptation auf Emotionen in der Stimme, aber nicht zwingend auf das Erkennen der Emotionen selbst, hat (Nussbaum et al., 2022). Nussbaum et al. (2022) haben hierzu ein Adaptationsparadigma genutzt, um die Zusammenhänge zwischen F0, Klangfarbe und den Emotionen Wut und Angst zu erforschen. Mit dem Adaptationsparadigma kann folglich die akustische Zusammensetzung der Emotionen und anderer Informationen in der Stimme erforscht werden.

In der gleichen Thematik könnte man die Rolle des Hörers mit einbringen. Wer nimmt Männer eher als wütend beziehungsweise Frauen als ängstlich wahr, weibliche oder männliche Hörer? Dies könnte ebenfalls mit einem Adaptationsparadigma untersucht werden, wobei in der Analyse eine getrennte Auswertung der weiblichen und männlichen ProbandInnen stattfinden würde. Auf der Seite des Hörers liegt ebenfalls die Frage, auf welchen Parametern die Verwechslung von Emotionen beruht und

welche Rolle das Geschlecht dabei spielt. Man weiß bereits, dass manche Emotionen schneller verwechselt werden als andere, sodass es bei der Unterscheidung von zum Beispiel Traurigkeit und Zärtlichkeit schneller zu Fehlern kommt als bei der Zuordnung von Traurigkeit und Freude (Juslin & Laukka, 2003). Interessant wäre hier der Einfluss des Geschlechts. Wenn Frauen tatsächlich in der Lage wären, Emotionen besser auszudrücken, dann dürften die Emotionen in weiblichen Stimmen auch besser erkannt werden als in männlichen. Gibt es Emotionen, die bei Frauen eher verwechselt werden als bei Männern?

Außerdem sollten die Ergebnisse in die Grundlagenforschung mit eingebracht werden. Es wäre interessant zu untersuchen, welche zusätzlichen Informationen Einfluss haben können auf die Wahrnehmung von emotionalen Stimmen. So könnte nach dem gleichen Prinzip dieser Studie untersucht werden, ob eine spezifische Adaptation abhängig von Alter, verschiedenen Persönlichkeitsmerkmalen oder sogar Identität stattfindet, da es laut Vermutungen möglich ist, Persönlichkeitseigenschaften in der Stimme zu erkennen (Skuk & Schweinberger, 2014). Die Frage bleibt, ob diese Eigenschaften so verarbeitet werden, dass es zu einem spezifischen Adaptationseffekt kommt. Neuere Erkenntnisse zeigen, dass Kontexte in der Wahrnehmung von Emotionen eine Rolle spielen, auch bei jenen Emotionen, die nicht über die Stimme wahrgenommen werden (Frühholz & Schweinberger, 2021). Diese Erkenntnisse gehen in eine ähnliche Richtung, in der die genauen Einflüsse auf die Wahrnehmung von Emotionen mit Hilfe des Adaptationsparadigmas genauer erforscht werden könnten. Darüber könnten Ansätze geliefert werden, wie sich die neurologische Verarbeitung bei der Wahrnehmung von Stimmen und insbesondere der Wahrnehmung von Emotionen, aber auch anderen Informationen, in der Stimme zusammensetzt. Diese stellen spannende Fragen dar für die zukünftige Erforschung der Stimme und der Informationen, die sie enthält.

### **Zusammenfassung**

Mithilfe der Online-Studie inklusive Nutzung des Adaptationseffekt konnte der systematische Einfluss des Geschlechts bei der Wahrnehmung von Emotionen untersucht werden und die Fragestellung beantwortet werden, ob es zu einem geschlechterspezifischen Adaptationseffekt kommt. Es fand sich eine Interaktion zwischen dem Geschlecht der Stimuli und der Wahrnehmung der Emotionen, sodass die Daten in beiden Adaptationstypen einen geschlechterspezifischen Nacheffekt der Adaptation aufweisen. Damit sind die hier erhobenen Daten deckungsgleich mit dem



Ergebnismuster von Bestelmeyer et al. (2010a) für emotionale Gesichter in der Modalität der Sprache. Des Weiteren finden sich auch ohne Adaptation Unterschiede in der Wahrnehmung zwischen männlichen und weiblichen Stimuli. Diese Unterschiede deuten entweder darauf hin, dass die Geschlechter Emotionen unterschiedlich gut ausdrücken können, oder dass Frauen in ihrer sozialen Rolle mit anderen Emotionen verbunden werden als Männer, wodurch es zu Unterschieden in der Wahrnehmung der Emotionen abhängig vom Geschlecht kommen könnte. Diese Daten belegen somit überzeugend, dass die Wahrnehmung des Geschlechts und die Wahrnehmung der Emotion in Stimmen interagieren.

## Abbildungsverzeichnis

Bestelmeyer, P. E. G, Rouger, J., DeBruine, L. M & Belin, P. (2010b). Auditory adaptation in vocal affect perception. *Cognition*, 117(2), S.220.

## Literaturverzeichnis

- Bachorowski, J.-A. (1999). Vocal Expression and Perception of Emotion. *Current Directions in Psychological Science*, 8(2), 53–57. <https://doi.org/10.1111/1467-8721.00013>
- Banse, R. & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of personality and social psychology*, 70(3), 614–636. <https://doi.org/10.1037/0022-3514.70.3.614>
- Bäumel, K.-H. (1991). Farbadaption bei einem Gelblich-Bläulich-System: Prüfung einer einfachen numerischen Relation. *Zeitschrift für Psychologie*, 199, 251–265. <https://epub.uni-regensburg.de/9025/>
- Belin, P., Bestelmeyer, P. E. G., Latinus, M. & Watson, R. (2011). Understanding voice perception. *British journal of psychology (London, England : 1953)*, 102(4), 711–725. <https://doi.org/10.1111/j.2044-8295.2011.02041.x>.
- Bestelmeyer, Jones, DeBruine, Little & Welling (2010a). Face aftereffects suggest interdependent processing of expression and sex and of expression and race. *Visual Cognition*, 18(2), 255–274. <https://doi.org/10.1080/13506280802708024>
- Bestelmeyer, Rouger, DeBruine & Belin (2010b). Auditory adaptation in vocal affect perception. *Cognition*, 117(2), 217–223. <https://doi.org/10.1016/j.cognition.2010.08.008>
- Bestelmeyer, P. E. G. & Mühl, C. (2021). Individual differences in voice adaptability are specifically linked to voice perception skill. *Cognition*, 210, 104582. <https://doi.org/10.1016/j.cognition.2021.104582>
- Brody, L. R. (2000). The socialization of gender differences in emotional expression: Display rules, infant temperament, and differentiation. In *Gender and Emotion* (1. Aufl., S. 24–47). Cambridge University Press. <https://doi.org/10.1017/cbo9780511628191.003>
- Brück, C., Kreifelts, B. & Wildgruber, D. (2011). Emotional voices in context: a neurobiological model of multimodal affective information processing. *Physics of life reviews*, 8(4), 383–403. <https://doi.org/10.1016/j.plrev.2011.10.002>.
- Bülthoff, H. H., Fahle, M., Gegenfurtner, K. R., Mallot H.A. (Hrsg.). (1998). *Visuelle Wahrnehmung: Beiträge zur 1. Tübinger Wahrnehmungskonferenz*. Knirsch. [https://pure.mpg.de/rest/items/item\\_1793991/component/file\\_3175784/content](https://pure.mpg.de/rest/items/item_1793991/component/file_3175784/content)
- Caldwell, A., Lakens, D. & Parlett-Pelleriti, C. (2019). *Power analysis with Superpower*. Self Published on GitHub, <https://aaroncaldwell.us/SuperpowerBook>. <https://scholar.google.de/citations?user=p8vbenoaaaaj&hl=de&oi=sra>
- Darwin, C. (1872). *The Expression of the Emotions in Man and Animals by Charles Darwin*. Murray.

- Ekman, P. (2016). What Scientists Who Study Emotion Agree About. *Perspectives on psychological science : a journal of the Association for Psychological Science*, 11(1), 31–34. <https://doi.org/10.1177/1745691615596992>
- Ekman, P. & Cordaro, D. (2011). What is Meant by Calling Emotions Basic. *Emotion Review*, 3(4), 364–370. <https://doi.org/10.1177/1754073911410740>
- Fant, G. (1971). *Acoustic Theory of Speech Production: With Calculations based on X-Ray Studies of Russian Articulations. Description and Analysis of Contemporary Standard Russian: Bd. 2*. De Gruyter. <https://books.google.de/books?id=qa-AUPdWg6sC>
- Frühholz, S. & Schweinberger, S. R. (2021). Nonverbal auditory communication - Evidence for integrated neural systems for voice signal production and perception. *Progress in Neurobiology*, 199, 101948. <https://doi.org/10.1016/j.pneurobio.2020.101948>
- Heller, D., Krummenacher, J., Bösch, U. (1998). Der Einfluß mehrerer Bewegungsrichtungen auf den Bewegungsnacheffekt. In Bülthoff, H. H., Fahle, M., Gegenfurtner, K. R., Mallot H.A. (Hrsg.), *Visuelle Wahrnehmung: Beiträge zur 1. Tübinger Wahrnehmungskonferenz* (S. 58). Knirsch.
- Juslin, P. N. & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychological Bulletin*, 129(5), 770–814. <https://doi.org/10.1037/0033-2909.129.5.770>
- Kawahara, Morise, Banno & Skuk, V. G. (2013). Temporally variable multi-aspect N-way morphing based on interference-free speech representations. In *2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*. IEEE. <https://doi.org/10.1109/apsipa.2013.6694355>
- Kawahara, H [H.], Morise, M., Takahashi, T., Nisimura, R., Irino, T. & Banno, H. (2008). Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing: Las Vegas, NV, 31 March - 4 April 2008* (S. 3933–3936). IEEE Service Center. <https://doi.org/10.1109/ICASSP.2008.4518514>
- Kiese-Himmel, C. (2016). *Körperinstrument Stimme*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-49648-0>
- Kreiman, J. & Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. Wiley-Blackwell.
- Lima, C. F. & Castro, S. L. (2011). Speaking to the trained ear: musical expertise enhances the recognition of emotions in speech prosody. *Emotion (Washington, D.C.)*, 11(5), 1021–1031. <https://doi.org/10.1037/a0024521>
- Nussbaum, C., Eiff, C. I. von, Skuk, V. G. & Schweinberger, S. R. (2022). *Vocal emotion adaptation aftereffects within and across speaker genders: Roles of timbre and fundamental frequency*. <https://doi.org/10.31234/osf.io/zcypa>
- Schandry, R. (2016). *Biologische Psychologie: Mit Online-Material* (4., überarbeitete Auflage). Beltz. <http://www.beltz.de/de/nc/verlagsgruppe-beltz/gesamtprogramm.html?isbn=978-3-621-28182-9>
- Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99(2), 143–165. <https://doi.org/10.1037/0033-2909.99.2.143>

- Schweinberger, S. R., Casper, C., Hauthal, N., Kaufmann, J. M., Kawahara, H [Hideki], Kloth, N., Robertson, D. M. C., Simpson, A. P. & Zäske, R. (2008). Auditory adaptation in voice perception. *Current biology : CB*, 18(9), 684–688. <https://doi.org/10.1016/j.cub.2008.04.015>.
- Schweinberger, S. R., Kawahara, H [Hideki], Simpson, A. P., Skuk, V. G. & Zäske, R. (2014). Speaker perception. *Wiley interdisciplinary reviews. Cognitive science*, 5(1), 15–25. <https://doi.org/10.1002/wcs.1261>
- Skuk, V. G. & Schweinberger, S. R. (2013). Adaptation aftereffects in vocal emotion perception elicited by expressive faces and voices. *PloS one*, 8(11), e81691. <https://doi.org/10.1371/journal.pone.0081691>
- Skuk, V. G. & Schweinberger, S. R. (2014). Influences of Fundamental Frequency, Formant Frequencies, Aperiodicity, and Spectrum Level on the Perception of Voice Gender. *Journal of Speech, Language, and Hearing Research*, 57(1), 285–296. [https://doi.org/10.1044/1092-4388\(2013/12-0314\)](https://doi.org/10.1044/1092-4388(2013/12-0314))
- Stoet, G. (2010). PsyToolkit: a software package for programming psychological experiments using Linux. *Behavior Research Methods*, 42(4), 1096–1104. <https://doi.org/10.3758/BRM.42.4.1096>
- Stoet, G. (2017). PsyToolkit. *Teaching of Psychology*, 44(1), 24–31. <https://doi.org/10.1177/0098628316677643>
- Uljarevic, M. & Hamilton, A. (2013). Recognition of emotions in autism: a formal meta-analysis. *Journal of Autism and Developmental Disorders*, 43(7), 1517–1526. <https://doi.org/10.1007/s10803-012-1695-5>
- Watzlawick, P., Bavelas, J. B. & Jackson, D. D. (2017). *Menschliche Kommunikation: Formen, Störungen, Paradoxien* (13. Aufl.). *Klassiker der Psychologie*. Hogrefe. <https://doi.org/10.1024/85745-000>
- Webster, M. A., Kaping, D., Mizokami, Y. & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature*, 428(6982), 557–561. <https://doi.org/10.1038/nature02420>
- Young, A. W., Frühholz, S. & Schweinberger, S. R. (2020). Face and Voice Perception: Understanding Commonalities and Differences. *Trends in cognitive sciences*, 24(5), 398–410. <https://doi.org/10.1016/j.tics.2020.02.001>.
- Zäske, R., Skuk, V. G., Kaufmann, J. M. & Schweinberger, S. R. (2013). Perceiving vocal age and gender: an adaptation approach. *Acta psychologica*, 144(3), 583–593. <https://doi.org/10.1016/j.actpsy.2013.09.009>

# Anhang

Darstellung der Daten der einzelnen ProbandInnen (N=33)

Anhang Abbildung 1: Adaptationsblock



Anhang Abbildung 1. Darstellung der Ergebnisse des Adaptationsblocks getrennt nach ProbandIn und Adaptationstyp für jeden Probanden einzeln. Finaler Datensatz N =33.

## VII



## **Instruktionen für die ProbandInnen zum Ablauf der Studie**

### **Instruktion 1: Beginn der Studie**

*Herzlich Willkommen!*

*Herzlich Willkommen zu unserem Hörexperiment zum Thema Emotionen in der menschlichen Stimme!*

*Alle weiteren Instruktionen finden Sie auf den folgenden Seiten.*

*Um zur nächsten Seite zu gelangen, drücken Sie die Leertaste.*

*\* Weiter durch Drücken der Leertaste \**

### **Instruktion 2: Beginn der Studie**

*Bitte machen Sie es sich bequem und stellen Sie sicher, dass sich in Ihrem Umfeld nichts mehr befindet, was Sie während des Experiments ablenken könnte.*

*Zu diesem Zweck schalten Sie bitte Ihr Handy und andere mögliche Störquellen lautlos und legen Sie diese für das Experiment weg.*

*Nutzen Sie außerdem bitte Kopfhörer und eine Tastatur für das Experiment.*

*Auf den nächsten Seiten beschreiben wir Ihnen den Ablauf und Ihre Aufgabe.*

*\* Weiter durch Drücken der Leertaste \**

### **Instruktion 3: Instruktion Baselineblock (1)**

#### **Aufgabenbeschreibung**

*Das Experiment besteht aus mehreren Teilen.*

*Vor jedem dieser Teile gibt es einen kurzen Übungsblock. Während des Experiments erhalten Sie in regelmäßigen Abständen die Gelegenheit, Pausen zu machen.*

*Vor jedem dieser Blöcke erhalten Sie Instruktionen zu Ihrer jeweiligen Aufgabe.*

*Daher bitten wir Sie, alle folgenden Instruktionen aufmerksam zu lesen.*

*\* Weiter durch Drücken der Leertaste \**

#### Instruktion 4: Instruktion Baselineblock (2)

##### **Aufgabenbeschreibung Teil 1**

*Im jedem Durchgang werden Sie zunächst ein grünes Fixationskreuz (+) sehen.  
Sobald es durch ein grünes Fragezeichen (?) ersetzt wird, wird Ihnen die jeweilige  
Tonaufnahme vorgespielt.*

*Bei jeder Tonaufnahme handelt es sich um jeweils ein Pseudowort, also kein tatsächlich  
im deutschen Sprachraum existierendes Wort.*

*Ihre Aufgabe besteht darin, zu entscheiden, ob die jeweils gespielte Tonaufnahme  
**wütend** oder **ängstlich** klingt und das per Tastendruck anzugeben.*

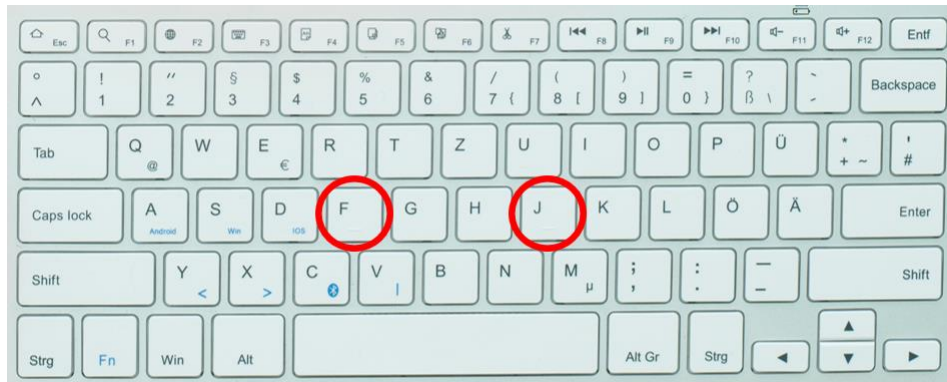
*Diese Entscheidung kann je nach Aufnahme unterschiedlich schwer sein. Bitte  
entscheiden Sie sich in jedem Durchgang für die Option, zu der Sie jeweils tendieren.  
Es geht dabei um Ihren **persönlichen Eindruck**, es gibt keine richtigen oder falschen  
Antworten.*

*\* Weiter durch Drücken der Leertaste \**

#### Instruktion 5: Instruktion Baselineblock (3)

##### **Aufgabenstellung**

*Geben Sie per Tastendruck von F oder J an, ob Sie die Stimme als ängstlich oder als  
wütend einschätzen würden:*



*F = ängstlich*

*J = wütend*

*Es folgt jetzt ein kurzer Übungsblock, damit Sie dem Ablauf vertraut werden.  
Bitte setzen Sie die Ihre **Kopfhörer** auf und legen Sie die Zeigefinger locker auf die  
angegebenen Tasten.*

*\* Weiter durch Drücken der Leertaste \**



#### Instruktion 6: Übungsblock Teil 1 Anfang

##### **Übung**

*Es folgt zunächst ein kurzer Übungsblock.*

*Bitte beachten Sie: jede Stimme wird nur **einmal** vorgespielt.*

*Bitte versuchen Sie, sie aufmerksam anzuhören.*

*Ordnen Sie die Stimmen den Emotionen per Tastendruck zu.*

*Noch einmal zur Erinnerung:*

*F = ängstlich*

*J = wütend*

*Durch Drücken der Leertaste starten Sie den Übungsdurchlauf.*

*\* Weiter durch Drücken der Leertaste \**

#### Instruktion 7: Übungsblock Teil 1 Ende

##### **Übung geschafft!**

*Der Übungsdurchlauf ist nun beendet.*

*Es folgt nun der erste Teil des eigentlichen Experiments. Bitte konzentrieren Sie sich auf die Ihnen dargebotenen Tonaufnahmen.*

*Sie werden in regelmäßigen Abständen die Möglichkeit bekommen, Pausen zu machen.*

*Noch einmal zur Erinnerung:*

*F = ängstlich*

*J = wütend*

*Bitte legen Sie nun ihre beiden Zeigefinger locker auf die Tasten F und J.*

*\* Weiter durch Drücken der Leertaste \**

#### Instruktion 8: Baselineblock Ende

##### **Erster Teil geschafft!**

*Super! 😊*

*Der erste Teil des Experiments ist beendet.*

*Im Folgenden beginnt der Haupttest.*

*\* Weiter durch Drücken der Leertaste \**

Instruktion 9: Instruktion Adaptationsblock (1)

**Aufgabenbeschreibung Teil 2**

*Dieser Teil des Experiments besteht aus verschiedenen Blöcken:*

*Es erscheint zuerst ein rotes Fixationskreuz (+) und dann werden hintereinander mehrere Stimmen abgespielt. Bitte hören Sie diese aufmerksam an!*

*Danach erscheinen auf dem Bildschirm für 3 Sekunden die Worte:*

*Jetzt geht es los!*

*Danach kommt die **Antwortphase**.*

*\* Weiter durch Drücken der Leertaste \**

Instruktion 10: Instruktion Adaptationsblock (2)

**Aufgabenbeschreibung Teil 2**

*In der Antwortphase werden Sie in jedem Durchgang zunächst ein **grünes** Fixationskreuz (+) sehen.*

*Sobald es durch ein grünes Fragezeichen (?) ersetzt wird, wird Ihnen die jeweilige Stimme vorgespielt werden.*

*Bitte bewerten Sie diese Stimme dann per Tastendruck, ob die Stimme für Sie ängstlich (F) oder wütend (J) klingt.*

**ACHTUNG:** *alle 4 Durchgänge erscheint noch einmal ein **rotes** Fixationskreuz (+) und es werden zwei Stimmen vorgespielt. Bitte hören Sie diese aufmerksam an. Bewerten müssen Sie diese nicht!*

*Danach kommt wieder ein grünes Fixationskreuz (+) und es geht weiter mit der Beurteilung per Tastendruck.*

*\* Weiter durch Drücken der Leertaste \**

## Instruktion 11: Übungsblock Teil 2 Anfang

### **Übungsblock**

*Nun folgt zunächst ein Übungsblock.*

*Noch einmal zur Erinnerung:*

*ängstlich = F*

*wütend = J*

*Es kommt zuerst die Vorlaufphase in welcher Sie NICHT antworten müssen. Sie erkennen diese an dem roten Fixationskreuz (+).*

*Nach den Worten „Jetzt geht es los!“ kommt dann die Antwortphase mit einem grünen Fixationskreuz (+). Nachdem Sie 4 mal bewertet haben erscheint ein rotes Fixationskreuz (+), bei dem Sie nichts bewerten müssen.*

*\* Weiter durch Drücken der Leertaste \**

## Instruktion 12: Übungsblock Teil 2 Ende

### **Übung geschafft!**

*Super!*

*Der Übungsblock ist beendet.*

*Im Folgenden beginnt Teil 2.*

*Dieser gliedert sich in mehrere kurze Blöcke.*

*Sie haben nach jedem Block die Möglichkeit, eine Pause zu machen.*

*\* Weiter durch Drücken der Leertaste \**

## Instruktion 13: Adaptationsblock Beginn

### **Teil 2**

*Noch einmal zur Erinnerung:*

*F = ängstlich*

*J = wütend*

*Es kommt zuerst die Vorlaufphase in welcher Sie nicht antworten müssen. Sie erkennen diese an dem roten Fixationskreuz (+).*

*Nach den Worten „Jetzt geht es los!“ kommt dann die Antwortphase mit einem grünen Fixationskreuz (+).*

*\* Weiter durch Drücken der Leertaste \**

#### Instruktion 14: Adaptationsblock Beginn

##### **Teil 2**

*Noch einmal zur Erinnerung:*

*F = ängstlich*

*J = wütend*

*Es kommt zuerst die Vorlaufphase in welcher Sie nicht antworten müssen. Sie erkennen diese an dem roten Fixationskreuz (+).*

*Nach den Worten „Jetzt geht es los!“ kommt dann die Antwortphase mit einem grünen Fixationskreuz (+).*

**ACHTUNG:** *alle 4 Durchgänge erscheint noch einmal ein **rotes** Fixationskreuz (+) und es werden zwei Stimmen vorgespielt. Bitte hören Sie diese aufmerksam an.*

*Danach kommt wieder ein grünes Fixationskreuz (+) und es geht weiter mit der Antwortphase.*

*\* Weiter durch Drücken der Leertaste \**

#### Instruktion 15: Pausen

***Pause***

***Super!***

***Sie haben wieder einen Block geschafft. ☺***

*Sie haben nun die Gelegenheit, eine kurze Pause zu machen und sich zu entspannen.*

*Wenn Sie sich bereit fühlen, weiterzumachen, drücken Sie bitte die Leertaste.*

*\* Weiter durch Drücken der Leertaste \**

#### Instruktion 16: Experiment Ende

*Das Experiment haben Sie erfolgreich beendet.*

*Vielen Dank für Ihre Teilnahme!*



*\* Weiter durch Drücken der Leertaste \**

### ***Eigenständigkeitserklärung***

Ich erkläre, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Hilfsmittel und Quellen angefertigt habe. Insbesondere versichere ich, dass ich alle wörtlichen Zitate und sinngemäßen Übernahmen aus anderen Werken als solche kenntlich gemacht habe.

Außerdem versichere ich, dass die vorliegende Arbeit die ethischen Richtlinien für Forschung (§7.3, Berufsethische Richtlinien) der Deutschen Gesellschaft für Psychologie (DGPs) erfüllt.

---

Ort, Datum

---

Unterschrift