

## AKUSTISCHE KORRELATE WAHrgENOMMENER PERSÖNLICHKEITSMERKMALE UND STIMMATTRAKTIVITÄT

*Daniel Duran, Natalie Lewandowski, Jagoda Bruni, Antje Schweitzer*

*Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart  
daniel.duran@ims.uni-stuttgart.de*

**Kurzfassung:** Wir präsentieren eine Studie über akustische Korrelate wahrgenommener Persönlichkeitsmerkmale und Attraktivität in Dialogsituationen. Wir verwenden Sprachaufnahmen aus der Erweiterung der GECO 2 Datenbank. Diese enthält insgesamt 60 spontane Dialoge zwischen 10 männlichen und 10 weiblichen Sprechern in verschiedenen Paarungen. Unmittelbar nach den Dialogen gaben alle Sprecher in einem Fragebogen subjektive Bewertungen ihrer Gesprächspartner ab. Dazu gehören Fragen nach der sozialen Attraktivität, der wahrgenommenen Kompetenz, Intelligenz oder Überlegenheit im Dialog. Für jeden Sprecher liegen mehrere derartige Bewertungen aus verschiedenen Dialogen vor. Wir untersuchen in dieser Studie, wie akustische Parameter, die laut Literatur mit der Wahrnehmung von Stimmattraktivität zusammenhängen, mit den gegenseitigen Bewertungen der Dialogpartner korrelieren. Wir extrahieren dafür Daten für Intensität, die Stimmgrundfrequenz sowie verschiedene Stimmqualitätsparameter aus den Sprachaufnahmen und werten diese statistisch aus.

### 1 Einführung

In zahlreichen Studien wurde beobachtet, dass emotionale Zustände sich in den akustischen Merkmalen der gesprochenen Sprache niederschlagen [1, 2, 3, 4, 5, u.a.]. Neuere Studien stellen z.B. fest, dass Variationen in der Stimmgrundfrequenz (F0) sowie der Intensität und der Sprechrate die Wahrnehmung von Emotionen beeinflussen [4, 6, 7, u.a.]. Traurigkeit zum Beispiel schlägt sich in einem geringen F0-Umfang, fallenden Pitchkonturen und einer geringeren Sprechrate aus; Angst hingegen in einer hohen Stimmgrundfrequenz, abrupten Änderungen in der F0-Variabilität und einer höheren Sprechrate. Ähnliche Muster wurden für Freude beobachtet. Langeweile wiederum zeigt sich in einer geringeren Variabilität [1, 2, 3, 4].

Auch die wahrgenommene Persönlichkeit wird von Merkmalen der Stimme beeinflusst. Schon der Begriff Persönlichkeit, vom lateinischen *personare* (durchtönen, widerhallen) wird auf die Mundöffnung in der Maske eines Schauspielers zurückgeführt [8] und bezieht sich somit auf die Stimme des Sprechers. Bereits antike Griechen und Römer scheinen sich mit emotionalen Zuständen in der gesprochenen Sprache befasst zu haben. Dabei wurde eine tiefe und angespannte Stimme mit Mut gleichgesetzt, eine hohe und schwache hingegen mit Feigheit [5]. Aronovitch [9] stellt fest, dass männliche Stimmen mit einer höheren Intensitätsvarianz als extrovertiert oder wagemutig wahrgenommen werden. Hingegen werden Stimmen erwachsener Sprecher mit einer höheren Stimmgrundfrequenz, die als kindlich wahrgenommen werden, mit einer weniger mächtigen dafür aber mit einer wärmeren und zugänglicheren Persönlichkeit assoziiert, möglicherweise auch mit geringerem Durchsetzungsvermögen, weniger Kompetenz und größerer Verwundbarkeit [10, 5].

Hinsichtlich der Wahrnehmung sozialer Eigenschaften findet Weiss [11] einen Zusammenhang zwischen Sympathie und folgenden Parametern: niedrigere Grundfrequenz (für beide Geschlechter), erhöhtes Sprechtempo bei weiblichen Sprechern, geringere Varianz in der Artikulationsrate bei weiblichen Sprechern, sowie geringere Varianz in der Intensität bei männlichen Sprechern. Neben der Varianz in der Stimmgrundfrequenz findet Weiss [11] auch Hinweise dafür, dass Diversifikation in der Sprechrate zur Wahrnehmung verschiedener sozialer Eigenschaften führt, wobei eine schnellere Sprechweise einer extrovertierten und kompetenten Persönlichkeit entsprechen und häufigere Pausen einer introvertierten und weniger sympathischen Persönlichkeit entsprechen.

Die Wahrnehmung von Stimmattraktivität wiederum wird zumindest teilweise von der wahrgenommenen Persönlichkeit beeinflusst. Kreiman und Sidtis [5] stellen fest, dass Menschen mit attraktiveren Stimmen als mächtiger, kompetenter, dominanter, erfolgreicher und sympathischer wahrgenommen werden als jene, deren Stimmen als weniger attraktiv bewertet werden. In diesem Zusammenhang wird eine attraktive Stimme mit einer entspannteren, selbstsicheren Persönlichkeit assoziiert. Weitere Belege für eine Abhängigkeit von Stimmattraktivität und wahrgenommener Persönlichkeit liefern Zuckerman und Driver [10]. Sie führten Perzeptionsstudien durch, um den Zusammenhang zwischen Stimmattraktivität und verschiedenen wahrgenommenen Persönlichkeitsmerkmalen zu untersuchen. Die Autoren stellen fest, dass Sprecher, deren Stimmen als attraktiver bewertet wurden, insgesamt positiver als andere bewertet wurden. Von zwei Teilnehmergruppen von Hörern, von denen eine nur die Persönlichkeit der Sprecher bewerten sollte und die andere nur die Stimmattraktivität gaben die Teilnehmer an, dass beide Aspekte zusammenhängen, auch wenn sie sich nicht der Tatsache bewusst waren, dass diese beiden Aspekte im Fokus der Studie standen. Insgesamt geben die Autoren an, dass die Stimmattraktivität am stärksten die Dominanz beeinflusst und die physische Attraktivität die Sympathie. Dies scheint darauf hinzudeuten, dass Stimmattraktivität sich nicht von sozialer Attraktivität trennen lässt.

Wie auch die Erkennung von Emotionen und Persönlichkeit ist die Fähigkeit, soziale und physische Attraktivität zu dekodieren, mit Eigenschaften der Stimme verbunden. So zeigen etwa Babel et al. [12], dass akustisch-phonetische Parameter die Wahrnehmung von sexuellem Dimorphismus, Gesundheit oder sozialer Attraktivität beeinflussen. Sie untersuchen den direkten Zusammenhang zwischen Stimmattraktivität und akustischen Parametern und stellen fest, dass akustische Merkmale, die mit der scheinbaren Größe des Sprechers, der scheinbaren Gesundheit und Jugendlichkeit sowie der sozialen Gruppenzugehörigkeit zusammenhängen, zur wahrgenommenen Stimmattraktivität beitragen. Babel et al. [12] vermuten, dass jedes akustische Merkmal, das das Geschlecht des Sprechers oder soziale Unterschiede signalisiert auch ein signifikanter Prädiktor für Stimmattraktivität sein kann. Auch D'Errico et al. [13] zeigen in ihrer Studie zu charismatischen Stimmen, dass die Stimmgrundfrequenz stark mit der Wahrnehmung sozialen und politischen Charismas<sup>1</sup> korreliert. Sprecher passen demnach ihre Stimmqualität an, um in in einer bestimmten Art wahrgenommen zu werden oder um bestimmte Konversationsziele zu erreichen. Desweiteren stellen Xu et al. [14] in einer Studie mit manipulierten und synthetischen Stimuli fest, dass männliche Hörer weibliche Stimmen mit stärkerer Behauchung, einer höheren F0 und einer breiteren Streuung der Formanten als attraktiver bewerten. Demgegenüber haben attraktivere männliche Stimmen eine niedrigere F0 und kompaktere Formantstrukturen.

Die meisten Studien verwenden relativ kontrollierte Stimuli für die Untersuchung der akustischen Parameter, wie z.B. gelesenes Material. Auch stammen Bewertungen hinsichtlich der wahrgenommenen Persönlichkeit bzw. hinsichtlich der konkreten Stimmattraktivität von un-

---

<sup>1</sup>D'Errico et al. [13] schließen Attraktivität in ihrer Definition von Charisma ein. Wir können hier nicht auf die Unterschiede eingehen.

beteiligten Hörern. Anders als in den meisten bisherigen Studien stammen in unserer Studie die Bewertungen vom jeweiligen Gesprächspartner und sind insofern maximal “authentisch”. Da eine attraktive Stimme stereotyp als Hinweis auf eine entspannte Persönlichkeit verstanden wird [15], ist es in dieser Hinsicht von Vorteil, dass die Dialogpartner aus unseren Daten sich vorher nicht kannten (siehe unten). Wir vermuten natürlich auch einen Zusammenhang zwischen Gesprächsinhalten und wahrgenommener Persönlichkeit; eine Analyse der Inhalte ist aber zum jetzigen Annotationsstand noch nicht möglich.

## 2 Methode

**Daten: GECO 2.** Wir nutzen für unsere Untersuchung 28<sup>2</sup> gemischt-geschlechtliche Dialoge der GECO 2-Datenbank [16]. Diese wurde erstellt, um phonetische Konvergenz in spontan-sprachlichen Dialogen zu untersuchen [17]. Die Sprecher kannten sich vorher nicht und saßen sich während der Dialoge in einem schallisolierten Aufnahmerraum durch eine Scheibe getrennt gegenüber. Die Dialoge dauern jeweils ca. 20 Minuten. Der GECO-Datensatz umfasst neben den Sprachaufnahmen auch subjektive Bewertungen der beteiligten Sprecher über die Dialogsituation, das eigene Befinden sowie über den jeweiligen Gesprächspartner. Es wurde nicht explizit nach Stimmattraktivität gefragt. Stattdessen bewerteten die Partner nach jedem Dialog, wie sympathisch, intelligent, freundlich, kompetent, erfolgreich, selbstsicher, sozial und locker sie sich auf einer Skala von 1 (“nein”) bis 5 (“ja”) gegenseitig wahrgenommen haben. Außerdem wurde unter anderem auch erfasst, ob das Gespräch als lustig empfunden wurde, auch dies auf einer Skala von 1 bis 5. Für jeden Sprecher liegen somit mehrere derartige Bewertungen aus verschiedenen Dialogen vor. Wir verwenden daher statt Stimmattraktivität näherungsweise die gegenseitigen Bewertungen für Sympathie, Freundlichkeit, sowie die Bewertung der Lustigkeit des Gesprächs. Diese stehen von allen in GECO erfassten Parameter am meisten in Zusammenhang mit wahrgenommener (Stimm-)attraktivität.

Die Vorverarbeitung der Dialogaufnahmen sowie alle akustischen Messungen wurden mit Praat [18] wie im Folgenden beschrieben durchgeführt.

**Turns.** Die Dialoge wurden folgendermaßen in sog. Turns, d.h. zusammenhängende Äußerungen eines Sprechers, segmentiert: Zunächst wurden Stilleintervalle identifiziert, innerhalb derer der lokale maximale Intensitätswert mindestens 45 dB unter dem globalen maximalen Intensitätswert lag, und deren Dauer mindestens 0,5 Sekunden betrug. Lautere Intervalle mit weniger als 50 ms Dauer wurden ebenfalls als Stille behandelt. Wir definieren also alle zwischen solchen Stilleintervallen auftretenden Äußerungen als Turns. Die Intensität wurde hierbei in Praat unter Berücksichtigung eines minimalen Pitchwerts berechnet, um eine Amplitudenmodulation der Intensität auszuschließen; hier wurde als minimaler Pitchwert bei Männern 80 Hz, bei Frauen 140 Hz angenommen. Für alle Aufnahmen wurde manuell überprüft, ob diese Defaults eine sinnvolle Segmentierung in Dialog-Turns ergaben und die Defaults ggf. entsprechend angepasst. Anschließend wurden alle Turns durch Anhören überprüft und ggf. einzelne Turngrenzen korrigiert. Die folgenden Schritte der Parameterextraktion basierten auf diesen handkorrigierten Segmentierung und wurden komplett automatisch durchgeführt.

**Grundfrequenz.** Zur Berechnung der Grundfrequenz (F0) mussten zunächst sinnvolle sprecherspezifische Minimal- und Maximalwerte bestimmt werden. Dabei folgten wir Hirst [19] und berechneten zunächst für jede Aufnahme eines Sprechers einmal die F0 unter Annahme eines F0-Ranges von 75 bis 600 Hz. Daraus berechneten wir ein individuelles F0-Minimum,

---

<sup>2</sup>Die Datenbank beinhaltet eigentlich 30 solcher Dialoge, sowie 30 gleichgeschlechtliche Dialoge. Allerdings waren aufgrund technischer Probleme zu diesem Zeitpunkt nur 28 dieser Dialoge auswertbar.

indem wir das 25%-Quantil der F0-Werte bestimmten und dies mit 0.75 multiplizierten, und ein individuelles F0-Maximum, indem wir das 75%-Quantil bestimmten und dieses mit 1.5 multiplizierten [19]. Anschließend wurde für jeden Turn einzeln die F0 noch einmal berechnet, und zwar mit den individuell bestimmten F0-Minima und F0-Maxima. Aus den so bestimmten F0-Konturen wurden anschließend wieder turnweise Minimum, Maximum und Mittelwert extrahiert.

**Intensität.** Die Intensität wurde turnweise direkt aus dem Sprachsignal berechnet, dabei floss das individuelle F0-Minimum mit ein.

**HNR.** Da die Harmonics-to-noise-Ratio normalerweise auf gehaltenen Vokalen berechnet wird, für die GECO-Daten aber noch keine Annotation auf Phonem-Ebene vorliegt, extrahierten wir für jeden Turn die maximale HNR, die unserer Einschätzung nach in vokalischen Intervallen vorliegen sollte. Zusätzlich wurde die Standardabweichung der max HNR berechnet, um die Größe der Schwankungen zu erfassen. Somit sollten die Werte automatisch die Harmonizität in Vokalen erfassen und somit eher einen Vergleich zwischen Turns bzw. zwischen Sprechern erlauben. Als Parameter floss das vorher etablierte Pitch-Minimum des Sprechers in die Berechnung mit ein.

**Jitter.** Zur Berechnung von Jitter wurden zunächst für jeden Turn die Anfänge der Grundperioden ermittelt (als PointProcess in Praat), wieder unter Berücksichtigung der individuellen F0-Minima und Maxima, und daraus dann der lokale Jitter berechnet, mit den von Praat vorgeschlagenen Default-Parametern. Dabei konnte aufgrund der noch nicht vorliegenden segmentalen Annotationen nicht systematisch dafür gesorgt werden, dass Jitter in vokalischen Bereichen berechnet wurde.

**Shimmer.** Shimmer wurde mithilfe des Signals und des oben genannten PointProcess berechnet. Auch hier wurden die von Praat vorgeschlagenen Defaults benutzt. Auch hier war es derzeit noch nicht möglich, sich systematisch auf vokalische Intervalle zu beschränken.

### 3 Ergebnisse

Die statistische Analyse wurde in R [20] mit den Bibliotheken *lme4* [21] und *car* [22] durchgeführt. Mit der Funktion *lmer* wurden lineare gemischte Modelle (*linear mixed models*) für jeden der oben genannten Parameter berechnet. Sprecher und Hörer (Bewerter) wurden als Random Intercepts integriert und die akustischen Parameter waren die jeweils abhängige Variable. Die Hörerbewertungen aus den Dialogen wurden als Prädiktoren für die akustischen Parameter verwendet. Die visuelle Inspektion der Residuenplots zeigte keine offensichtlichen Auffälligkeiten bzw. keine Heteroskedastizität. Um die Kolinearität der Faktoren zu beurteilen, wurden für alle Modelle die *vif scores*<sup>3</sup> der verwendeten Faktoren berechnet: Variablen, die im post-hoc Test für das jeweilige Modell einen *vif score* von über 4 hatten, wurden aus dem Modell entfernt aufgrund einer möglichen Kolinearität zwischen den Parametern. Die Modelle wurden untereinander mit der Funktion *anova* verglichen, um einen signifikanten Einfluss der unabhängigen Variablen auf die vorherzusagende Variable zu verifizieren. Für die statistische Analyse verwenden wir gemittelte Werte der extrahierten akustischen Parameter über einen gesamten Dialog. Die verschiedenen statistischen Modelle sind in Tabelle 1 zusammengefasst<sup>4</sup>.

Die im Fall der mittleren F0 im Modell verbliebenen Faktoren können der ersten Spalte der Tabelle 1 entnommen werden. Für die mittlere F0 finden wir einen signifikanten Effekt für das

<sup>3</sup><https://raw.githubusercontent.com/aufrank/R-hacks/master/mer-utils.R>

<sup>4</sup>Für die Darstellung wurde die R-Bibliothek *texreg* verwendet [23]

**Tabelle 1** – Überblick über die Modelle der verschiedenen akustischen Parameter (Spalten) und die für die diversen Faktoren (Zeilen) gefundenen Koeffizienten und deren Standardfehler. Ein Doppelpunkt steht für eine Interaktion der jeweiligen Variablen.

	mittl. F0	Intensität	F0-Umfang	Shimmer	Jitter	std. HNR
(Intercept)	222.1065*** (4.5250)	52.6869*** (0.5444)	487.1415*** (4.1589)	0.0916*** (0.0029)	0.0246*** (0.0010)	20.3181*** (1.3981)
lustig	0.9764 (2.9246)					
freundlich	0.6926 (2.4669)	-0.1700 (0.2929)		0.0007 (0.0008)		
SG	-98.2426*** (6.4146)	-0.5972 (0.7725)	51.6675*** (5.9361)	0.0261*** (0.0041)	0.0109*** (0.0014)	-5.5854** (1.9279)
lustig:SG	11.7953* (4.7050)					
freundlich:SG	-7.2133 (4.6058)	1.1230* (0.5127)				
sympathisch			5.0596* (2.0263)			
SG:freundlich				-0.0029* (0.0014)		
kompetent					-0.0011** (0.0004)	
kompetent:SG					-0.0001 (0.0006)	
erfolgreich						3.9212** (1.3758)
erfolgreich:SG						-4.5337** (1.7381)

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ;  $p$ -Werte ermittelt von *texreg*, für alle so markierten Faktoren gilt  $|t| > 2$ .

Geschlecht des Sprechers (SG) ( $t = -15,3$ ), der aufgrund der generell tieferen Männerstimmen natürlich nicht überraschend ist, sowie die Interaktion von SG mit dem Parameter *lustig* ( $t = 2,5$ ). Ein direkter Modellvergleich mittels anova ergab einen signifikanten Einfluss der Interaktion von *lustig* und SG ( $\chi^2 = 10,32$ ,  $p = 0,005749$ ), die einen Anstieg der mittleren F0 um 11,8Hz ( $\pm 4,7$  Hz Standardfehler) nach sich zog. Als Referenzlevel für SG galten hier und in allen weiteren Modellen die männlichen Sprecher.

Für die mittlere Intensität (zweite Spalte in Tabelle 1) zeigt die Interaktion der Parameter *Sprechergeschlecht* und *freundlich* einen signifikanten Effekt ( $t = 2,19$ ). Das Hinzufügen der Interaktion hat einen signifikanten Einfluss auf den Modellfit ( $\chi^2 = 4,57$  bei einem  $p = 0,03256$ ).

Der F0-Umfang zeigt signifikante Effekte für SG ( $t = 8,7$ ) und *sympathisch* ( $t = 2,5$ ). Das Hinzufügen des Parameters *sympathisch* hat laut anova einen signifikanten Einfluss auf den Fit des Modells ( $\chi^2 = 4,5652$ , bei  $p = 0,03263$ ) und bewirkt einen Anstieg des F0-Umfangs der Sprecher um 5.06 dB ( $\pm 2.03$  dB Standardfehler).

Das Modell für den mittleren Wert von *Shimmer* zeigt einen signifikanten Effekt für SG ( $t = 6,38$ ) sowie für die Interaktion von SG und *freundlich* ( $t = -2,16$ ). Die Miteinbeziehung der Interaktion zwischen SG und *freundlich* verbessert das Modell deutlich ( $\chi^2 = 3,9773$ ,  $p = 0,04612$ ). Mit steigendem Freundlichkeitswert verringert sich für männliche Sprecher der mittlere Shimmerwert um  $-0,00294$  (Standardfehler=  $\pm 0,00136$ ).

Das Modell für den mittleren Wert von *Jitter* zeigt den besten Fit, wenn neben dem Ge-

schlecht nur der Parameter *kompetent* berücksichtigt wird, mit einem signifikanten Effekt für *kompetent* ( $t = -2,9$ ) und SG ( $t = 7,85$ ). Ein unmittelbarer Vergleich des Akaike Informationskriteriums (AIC) und des Bayesianischen Informationskriteriums (BIC) von zwei Modellen (mit und ohne *kompetent*) bestätigte den besseren Fit des Modells mit der Variable *kompetent* (niedrigere AIC und BIC als das Vergleichsmodell).

Für die Standardabweichung der HNR Werte wurde ein lineares gemischtes Modell angepasst, das die Faktoren *Sprechergeschlecht* und *erfolgreich* berücksichtigt hat. Sowohl SG ( $t = -2,90$ ) als auch *erfolgreich* ( $t = 2,85$ ) und die Interaktion der beiden Parameter ( $t = -2,61$ ) beeinflussten die abhängige Variable. Im Modellvergleich mit anova schnitt das Modell mit Berücksichtigung von *erfolgreich* signifikant besser ab als ein reduziertes Modell ( $\chi^2 = 7,7841$  bei  $p = 0,0204$ ). Bei den erfolgreicher bewerteten männlichen Sprechern sehen wir eine um 4,534 reduzierte HNR Abweichung.

## 4 Diskussion und Ausblick

Wir können nicht viele der in der Literatur beschriebenen Parameter bestätigen. Dies mag daran liegen, dass wir es in dieser Studie mit spontansprachlichen Dialogen zu tun haben, die durch die beteiligten Dialogpartner selbst global bewertet wurden. Dabei dürften viele Faktoren eine Rolle spielen, die in dieser Studie unberücksichtigt bleiben mussten.

Unsere Ergebnisse zeigen, dass die mittlere F0 bei Männern um 11 Hz höher ist, wenn Dialoge als lustig bewertet wurden. Dies stimmt mit anderen Studien überein [9]. Es hat sich auch gezeigt, dass Sprecher, die als sympathischer wahrgenommen wurden, mit einer tieferen Stimme sprechen, wenn der Dialog als lustig empfunden wurde. Diese Beobachtung kann im Hinblick auf die Studie von Zuckerman und Miyake [24] interpretiert werden, welche zeigte, dass Stimmen, die als kompetent und sympathisch wahrgenommen werden, häufig mit einer deutlichen Artikulation, einem weniger schrillen Klang, geringerer Nasalität und guten Resonanzen zusammenhängt.

Frühere Studien [5, 25, 12] zeigen, dass die Bewertung von Stimmatraktivität sehr subjektiv und kulturabhängig ist. Wir verwenden Aufnahmen spontaner Sprache und die gegenseitigen Bewertungen der Gesprächspartner, die sich während des Dialogs auf engstem Raum in einem schallisolierten Raum direkt gegenüber saßen. Andere Studien verwenden kontrollierte Sprachaufnahmen (oft gelesener Sprache) und Bewertungen durch unbeteiligte Hörer basierend auf akustisch dargebotenen Stimuli. Dadurch ist ein direkter Vergleich unserer Ergebnisse mit anderen Studien zur Stimmatraktivität schwierig. Es fehlt insbesondere eine kultur- und kontext-unabhängige Definition eines univerellen Merkmalsinventars zur Erfassung von Stimmatraktivität, das die gesamte Bandbreite an relevanten akustischen und physiologischen Eigenschaften umfasst.

Unsere Ergebnisse sind ein erster Beitrag, subjektive Dialogbewertungen spontaner Sprache mit objektiven, akustischen Parametern in Verbindung zu bringen. Detailliertere akustisch-phonetische Analysen müssen durchgeführt werden, um z. B. den Einfluss der Formantstruktur oder anderer segmentaler Merkmale auf die Stimmatraktivität zu untersuchen. Im nächsten Schritt müssen Perzeptionsstudien durchgeführt werden, die explizite Bewertungen der Stimmatraktivität auf den selben Dialogdaten erheben. Damit ließe sich der Zusammenhang zwischen den globalen Bewertungen der beteiligten Gesprächspartner, der wahrgenommenen Stimmatraktivität durch unbeteiligte Hörer sowie verschiedener akustischer Merkmale untersuchen.

## Literatur

- [1] SCHERER, K. R.: *Vocal affect expression: A review and a model for future research. Psychological Bulletin*, 99(2), S. 143–165, 1986.
- [2] SCHERER, K.: *Vocal communication of emotion: A review of research paradigms. Speech Communication*, 40(1-2), S. 227–256, 2003.
- [3] MURRAY, I. R. und J. L. ARNOTT: *Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. The Journal of the Acoustical Society of America*, 93(2), S. 1097–1108, 1993.
- [4] JUSLIN, P. N. und P. LAUKKA: *Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. Emotion*, 1(4), S. 381–412, 2001.
- [5] KREIMAN, J. und D. SIDTIS: *Foundations of voice studies: an interdisciplinary approach to voice production and perception. Wiley-Blackwell, Chichester*, 2011.
- [6] BACHOROWSKI, J.-A. und M. J. OWREN: *Sounds of emotion: Production and perception of affect-related vocal acoustics. Annals of the New York Academy of Sciences*, 1000(1), S. 244–265, 2003.
- [7] BACHOROWSKI, J.-A. und M. J. OWREN: *Emotion-related vocal acoustics: Cue-configuration, dimensional, and affect-induction perspectives. In K. IZDEBSKI (Hrsg.), Emotions in the Human Voice, Volume I: Foundations*, S. 87–100. Plural Publishing, 2008.
- [8] KRAMER, E.: *Personality stereotypes in voice: A reconsideration of the data. The Journal of Social Psychology*, 62(2), S. 247–251, 1964.
- [9] ARONOVITCH, C. D.: *The Voice of Personality: Stereotyped Judgments and their Relation to Voice Quality and Sex of Speaker. The Journal of Social Psychology*, 99(2), S. 207–220, 1976.
- [10] ZUCKERMAN, M. und R. E. DRIVER: *What sounds beautiful is good: The vocal attractiveness stereotype. Journal of Nonverbal Behavior*, 13(2), S. 67–82, 1989.
- [11] WEISS, B.: *Prosodische Elemente vokaler Prosodie. In P. WAGNER (Hrsg.), Elektronische Sprachsignalverarbeitung 2013 – Tagungsband der 24. Konferenz*, S. 212–217. TUDpress, Dresden, 2013.
- [12] BABEL, M., G. MCGUIRE, und J. KING: *Towards a more nuanced view of vocal attractiveness. PLoS ONE*, 9(2), S. e88616, 2014.
- [13] D’ERRICO, F., R. SIGNORELLO, D. DEMOLIN, und I. POGGI: *The perception of charisma from voice: A cross-cultural study. In 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, S. 552–557. IEEE, 2013.
- [14] XU, Y., A. LEE, W.-L. WU, X. LIU, und P. BIRKHOLZ: *Human vocal attractiveness as signaled by body size projection. PLoS ONE*, 8(4), S. e62397, 2013.
- [15] ZUCKERMAN, M., H. HODGINS, und K. MIYAKE: *The vocal attractiveness stereotype: Replication and elaboration. Journal of Nonverbal Behavior*, 14(2), S. 97–112, 1990.

- [16] SCHWEITZER, A., N. LEWANDOWSKI, und D. DURAN: *Attention, please! Expanding the GECO database*. In THE SCOTTISH CONSORTIUM FOR ICPHS 2015 (Hrsg.), *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*. Glasgow, UK, 2015. Paper number 620.
- [17] LEWANDOWSKI, N.: *Talent in nonnative phonetic convergence*. Doctoral dissertation, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, 2012. doi:10.18419/opus-2858.
- [18] BOERSMA, P. und D. WEENINK: *Praat: doing phonetics by computer*. 2016. URL <http://www.praat.org/>. Version 6.0.
- [19] HIRST, D.: *A praat plugin for MOMEL and INTSINT with improved algorithms for modelling and coding intonation*. In *Proceedings of ICPhS XVI*, S. 1233–1236. 2007.
- [20] R CORE TEAM: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. URL <https://www.R-project.org/>.
- [21] BATES, D., M. MÄCHLER, B. BOLKER, und S. WALKER: *Fitting linear mixed-effects models using lme4*. *Journal of Statistical Software*, 67(1), S. 1–48, 2015.
- [22] FOX, J. und S. WEISBERG: *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, second edn., 2011. URL <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>.
- [23] LEIFELD, P.: *texreg: Conversion of statistical model output in R to L<sup>A</sup>T<sub>E</sub>X and HTML tables*. *Journal of Statistical Software*, 55(8), S. 1–24, 2013. URL <http://www.jstatsoft.org/v55/i08/>.
- [24] ZUCKERMAN, M. und K. MIYAKE: *The attractive voice: What makes it so?* *Journal of Nonverbal Behavior*, 17(2), S. 119–135, 1993.
- [25] SIGNORELLO, R. und D. DEMOLIN: *The physiological use of the charismatic voice in political speech*. In *INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association*, S. 987–991. ISCA Archive, Lyon, France, 2013.