Zur Wirkung menschlicher Stimme auf die wahrgenommene Sympathie – Einfluss der Stimmanregung –

Benjamin Weiss, Sebastian Möller, Tim Polzehl

Quality & Usability Labs, Deutsche Telekom Laboratories, TU Berlin
BWeiss@telekom.de

Zusammenfassung

Beurteilungen der zugeschriebenen Sympathie wurden für fünf weibliche und fünf männliche Sprecher aus der emoDB erfasst und zeigen eine signifikante Rangfolge unabhängig vom Geschlecht der Sprecher oder dem der Bewerter. Akustische Analysen der Parameter von Elektro-Glottogrammen (EGG) zeigen – getrennt für männliche und weibliche Sprecher – signifikante Korrelationen mit den subjektiven Urteilen. Allein aufgrund der EGG-Parametrisierungen können über 60% der Varianz in den Sympathiebewertungen erklärt werden. Die mögliche Bedeutung der Parameter für die Eindruckswirkung auf die Beurteiler wird abschließend diskutiert.

1 Einleitung

Prosodische Merkmale werden beim Zuhören ausgewertet, um allein aufgrund dieser akustischen Informationen *körperliche Charakteristika des Sprechenden* – etwa Geschlecht, Alter oder Emotion – mit zufriedenstellender Reliabilität zu schätzen [5]. Dies trifft jedoch nicht notwendigerweise für alle solche Merkmale zu – wie etwa Gewicht oder Körpergröße [11]. Beispielsweise waren weibliche Hörer zwar konsistent in ihren Bewertungen von Größe, Gewicht und Alter männlicher Stimmen, jedoch lagen sie bei der Körpergröße falsch [4]. Außerdem können Hörer aufgrund von Stimme und Sprechweise *Persönlichkeitsmerkmale* zuschreiben, z.B. die Extroversion-Introversion Opposition [24].

Laut *Fährmann* (1982), der Qualitätsaspekte von Stimme und Sprechweise in „habituelle“ (Stimmlage, Lautstärke, Klang, Stimmfülle), „individuelle“ (Tempo, Rhythmus, Akzentuierung, Artikulation) und „hinzutretende Qualitäten“ einteilt

(Sprechstil, Syntax, Wortwahl), spielt Stimmklang eine bedeutende Rolle bei der Zuschreibung und dem Ausdruck von Persönlichkeit [13]. Auch Aussprache, Intonation, Redefluss oder Pausensetzung kann die *Wirkung* auf den Eindruck auf und die *Bewertung* durch den Zuhörer beeinflussen [22, 19, 12, 3]. Für diese *Zuschreibung* von Sympathie kann nicht – wie etwa für die Persönlichkeitszuschreibung – auf theoretische und empirische Erkenntnisse aus der Psychologie aufgebaut werden (vgl. etwa [21]), da sich die Arbeiten zur sogenannten sozialen Attraktion verständlicherweise auf die interpersonellen Erfahrungen konzentrieren, nicht auf erste Eindrücke. Vielmehr beeinflusst Sympathie wiederum das Kommunikationsverhalten [16]. Dennoch ist das Zuschreiben von Sympathie im Alltag ein grundlegender Vorgang und wird auch regelmäßig mit der Stimme in Verbindung gebracht [1].

Beispielsweise wurde die Akzeptanz von Stimmen für den E-Learning Bereich untersucht [18]. In der Analyse von 15 Sprechern und 15 Sprecherinnen [2] zeigten sich eine tiefe Tonlage, eine „klare“ Stimme und ein gleichmäßige Stimmstärke (qualitativ bewertet) als angenehm. Anhand einer anschließenden Untersuchung fand *Ketzmerick* (2007) Korrelationen der Bewertungen mit akustischen Parametern, vor allem minimale Grundfrequenz und der Grundfrequenz Häufigkeitsverteilung.

Hörer bewerten insgesamt tiefere Stimmen besser [23]. Jedoch scheint dieses Ergebnis kontextabhängig zu sein, da auch das Gegenteil zutreffen kann [20]: Die höhere Stimmlage für einen Politiker wurde auf einem Parteitag besser bewertet, da die damit einhergehende Aktivität in dieser speziellen Situation angemessener war. Bei der Untersuchung von isoliert geäußerten Vokalen zeigte sich eine steigende Grundfrequenzkontur als Indikator für angenehme Männerstimmen für Frauen [4].

Nicht nur die Grundfrequenz, auch der zweite Formant beeinflusst die Bewertungen von Männerstimmen [17]. Die Autoren der Untersuchung empfehlen die Auswahl von dunklen Stimmen für TTS Systeme. Für die Anwendung in der Werbung führten höheres Sprechtempo, adäquate Pausensetzung und tiefere Grundfrequenz zu positiveren Bewertungen von Sprechern [7].

Die hier angeführten Ergebnisse illustrieren, dass bisher ausschließlich akustische Untersuchungen des Sprachsignals hinsichtlich Sympathie durchgeführt wurden. Insbesondere die Stimmlage (mittlere Tonhöhe der Grundfrequenz) wurde regelmäßig betrachtet. Akustische Analysen der Klangfarbe (Höreindruck des Schallspektrums) sind dagegen selten – zu ihrer Erfassung werden unter anderen auch qualitative Methoden in Form von Hörerbeschreibungen angewendet. Eben diese habituellen Qualitäten lassen sich jedoch auch in der Phonation wiederfinden, beispielsweise anhand der Anspannung bei der Vokalisierung oder der Fülle im Spektrum [10]. Noch fehlen genauere Erkenntnisse, welche Merkmale einen Menschen sympathisch klingen lassen. Es ist aber zu erwarten, dass siche-

re Sprecher, also nicht zu angespannte, jedoch kräftige Anregung sympathisch wirkt; ebenso eine klangvolle Stimme, also beispielsweise eine obertonreiche Anregung der Indifferenzlage. Aus diesem Grund wurden Laryngogramme (EGG) von Sprechern darauf untersucht, ob und welche Parameter der Glottissignale mit Bewertungen von Sympathie korrelieren. Im Folgenden wird dazu die Datenbasis vorgestellt, also Bewertungen der Sprecher, ihre Auswahl und Aufnahme der Laryngogramme, die akustischen Analysen und ihre Korrelation zu den subjektiven Bewertungen. Abschließend werden die gewonnenen Ergebnisse eingeordnet und diskutiert.

2 Material

Für ein grundlegendes Experiment wurden neun emotional neutral geäußerte Sätze von zehn Sprechern (5 weiblich, 5 männlich) aus der emoDB [6] ausgewählt. Die Sätze zeichnen sich durch neutrale Inhalte aus und bestehen aus ein bis zwei Intonationsphrasen. Dabei wurden die professionellen Sprecher/Schauspieler instruiert, natürlich und nicht übertrieben zu artikulieren.

Die Stimuli wurden von 20 Probanden auf einer 7-stufigen Skala auf ihren Eindruck von Sympathie bewertet (7 Frauen, 13 Männer, im Alter von 21–37 Jahren, M=27.7, SD=4.34). Für die Präsentation des Monosignals auf beiden Ohren diente ein Kopfhörer (Sennheiser HD 485). Die Stimuli des Satzes „a07“ wurden zum Training verwendet.

Wie andern Orts dargestellt, unterscheiden sich die mittleren Bewertungen der verwendeten Sätze nicht. Jedoch sind die Bewertungen für die Sprecher signifikant unterschiedlich, sodass eine Rangfolge von sympathischen bis unsympathischen Sprechern erstellt werden kann (vgl. [25]).

3 Durchführung

Um Informationen über die stimmliche Anregung zu erhalten, wurden abgesehen von Sprachsignalen auch Laryngogramme von den Sprechern aufgezeichnet (Laryngograph Ltd.). Dabei wird der elektrische Fluss der Glottis anhand von zwei seitlich angebrachten Elektroden erfasst. Die Amplitude des erfassten Signals beschreibt also die Stromstärke, nicht den Stimmlippenabstand oder den Luftdruck. Auch gibt die Amplitude nicht die Stimmintensität wider [14]. Statt dessen lässt sich erhöhte Lautstärke durch die Auswirkungen des erhöhten subglottalen Luftdrucks erfassen: Etwa ein flach abfallendes Quellspektrum.

Die Phonation bestimmt nicht nur die Stimmqualität bei der Lautbildung, sondern auch die Tonhöhe und Intonation. Einflussfaktoren auf die Sympathie könn-

ten einerseits statische Faktoren sein, wie die Anatomie des Kehlkopfes, jedoch auch dynamische, wie etwa die Erregung, die sich in der Höhe des Kehlkopfes und der Spannung der Muskeln mit ihren Veränderungen der beherrschenden Kräfte (mediale, longitudinale, adduktive). Beispielsweise führt eine schnellere Verschlussphase zu stärker ausgeprägten Impulsen mit stärkeren Obertönen des Anregungssignals.

So erfasst ein EGG Informationen zur Intonation, aber auch zur Behauchung, Gepresstheit [9] etc. Das Signal lässt sich für jede Periode in drei Phasen einteilen: die Öffnungsphase, die Schließphase und geschlossene Phase [8]. Typische Parametrisierungen sind die so erhaltene Grundfrequenz (*F0*) samt Streumaß (etwa der Standard Abweichung, *F0 SD*) und der Prozentsatz der Dauer geöffneter Stimmlippen zur Gesamtdauer (*Open Quotient*).

Zusätzlich zu diesen beiden wurden aus den Signalen der Sprecher folgende Parameter extrahiert und über die einzelnen Stimuli gemittelt: Tempo der Verschlussphase (*Derivative Peak*) als Amplitude des positiven Maximums der Ableitung, die mittlere *Amplitude der Ableitung* und dessen Standard Abweichung als Maße für die Dynamik der Stimmlippenbewegungen, sowie Maße zur Beschreibung des Spektrums des Anregungssignals (*Center of Gravity, Kurtosis, Skewness, Verhältnis der Energie zwischen hohen und tiefen Frequenzen* (cut-off frequency bei doppelter Grundfrequenz)) und das Harmonic-to-Noise Verhältnis. Die ermittelte Grundfrequenz dient dem Abgleich mit der akustischen Analyse [25]. Jitter und Shimmer der Grundfrequenz wurden ebenfalls erhoben. Das Maximum der Ableitung bestimmt den Beginn der geschlossenen Phase und wurde zur Messung der Grundfrequenz verwendet. Dieses Maximum und das lokale Minimum werden zur Bestimmung des *Open Quotient* herangezogen [15].

4 Ergebnisse

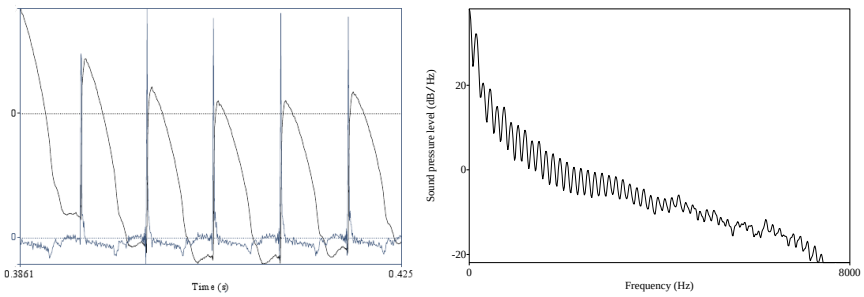

Abbildung 1: links: Ausschnitt aus dem EGG (schwarz) und der Ableitung (blau) eines männlichen Sprechers; rechts: dazugehörige Spektrum.

Die akustischen Maße aus dem Anregungssignal wurden nach Extraktion der stimmhaften Zeitabschnitte über jeden Stimulus gemittelt. Speziellere Analysen (einzelner vergleichbarer Vokale beispielsweise) wurden noch nicht vorgenommen. In Tabelle 1 sind diejenigen der erhobenen Parameter aufgeführt, die einen signifikanten Beitrag zu den gemittelten Hörerbewertungen der Sympathie leisten.

Tabelle 1: Signifikante Ergebnisse eines Multiplen Linearen Modells für die gemittelten Bewertungen für jeden Stimulus.

Parameter	Geschlecht	t-Wert (DF=76)	p
Shimmer	m	-2.44	0.0172*
CoG	m	-2.27	0.0258*
CoG	w	4.27	0.0001***
CoG SD	m	2.52	0.0140*
CoG SD	w	-3.33	0.0014**
dEGG peak	w	-2.44	0.0170*
Amplitude dEGG SD	w	2.10	0.0402*
Differenz vom mittl. Open Quotient	w	-2.14	0.0356*

Natürlich ergeben sich hier Interaktionseffekte mit dem Geschlecht, da die Grundfrequenz und das Spektrum von der Kehlkopfgröße anhängen. Obwohl keine Geschlechter-Unterschiede in den Bewertungen selbst auftraten, können für die Interaktionseffekte auch unterschiedliche Bewertungsmaßstäbe – erfasst über die verschiedenen Parameter und Parameterausprägungen – für Frauen und Männerstimmen verantwortlich sein, oder die Sprecherauswahl widerspiegeln. Beurteilungen von *Sympathie wurde positiver bewertet mit:*

- *geringerem Shimmer* (männliche Sprecher),
- *geringerem Center of Gravity* (männliche Sprecher),
- *höherem Center of Gravity* (weibliche Sprecher),
- *höherer Variabilität in dem Center of Gravity* (männliche Sprecher),
- *niedrigerer Variabilität in dem Center of Gravity* (weibliche Sprecher),
- *niedrigerem Derivative Peak zur Verschlussphase* (weibliche Sprecher),
- *höherer Variabilität in der Amplitude der Ableitung* (weibliche Sprecher)
- *durchschnittlicherem Open Quotient* (weibliche Sprecher).

Die Abweichung vom mittleren *Open Quotient* wurde nach Ansicht der Daten als Alternative zu *Open Quotient* hinzugezogen. Eine Regressionsanalyse mit den signifikanten Parametern erklärt 61% der Varianz ($R_{adjust} = .54$, $RSE = .55$, $p < .001$ und ist damit deutlich besser als ein vergleichbares Modell (auch sechs einbezogene Parameter) mit akustischen Parametern aus dem Sprachsignal ($R_{adjust} = .37$, $RSE = .64$, $p < .001$; vgl. [25]. Dies ist umso erstaunlicher, da die Sprachsignale derselben Aufnahmen abgesehen von den Parametern der Grundfrequenz

auch spektrale Parameter und das signifikante Sprechtempo enthalten und damit den Einfluss des Vokaltraktes zusätzlich erfassen.

5 Diskussion

Bereits die Analyse der stimmhaften Anteile des Anregungssignals reicht aus, um subjektive Bewertungen von Sympathie in einem zufriedenstellendem Maß zu erklären. Leider ist die Anzahl an verwendeten Sprechern zu gering, um die Resultate der Analyse zu generalisieren; insbesondere aufgrund der notwendigen Geschlechtertrennung. Deswegen wurde auch auf eine Darstellung der tatsächlichen Parameterausprägungen verzichtet. Jedoch ist dieses Ergebnis vielversprechend für weitere Untersuchungen. Zudem muss noch geklärt werden, ob sich die akustischen Parameter für beide Geschlechter normalisieren lassen, oder hier unterschiedliche Wirkmechanismen vorliegen, wie etwa durch die gegensätzlichen Ergebnisse für die Variabilität der Amplitude der Ableitung suggerieren.

In Ansätzen kann versucht werden, die Bedeutung der erfassten Parameter auf ihre Ursache und Eindruckswirkung beim Hörer zu interpretieren: Laut [10] weist das Anregungssignal fünf perzeptive Dimensionen auf: *Flow* (hier erfasst über „Derivative Peak (dEGG peak)“), *Tenseness* („Open Quotient“), *Loudness* („Energieverhältnis“), *Roughness* („F0-Jitter“ und „Shimmer“) und *Breathiness* („Harmonics-to-Noise Ratio“).

Bei Bestätigung der hier dargestellten Ergebnisse mit weiteren Sprechern würden positivere Bewertungen der Frauenstimmen mit durchschnittlicher Tenseness, vermindertem Flow (dEGG peak, ggf. als Zeichen sanfterer Anregung), lebendigerem Flow (Variabilität in der Amplitude der Ableitung), höherem CoG (als Zeichen einer helleren Stimme), und weniger Variabilität im „CoG“ einhergehen. Für die männlichen Stimmen träfe dies für geringere *Roughness* (Shimmer), niedrigerem CoG (als Zeichen einer dunkleren Stimme) und höherer Variabilität in dem „CoG“ zu.

Auffallend ist das Fehlen des Einflusses der Grundfrequenz, der für die Sprachsignale in [25] signifikant ist. Die anderen Parameter erklären die Varianz in den hier vorgestellten Daten besser. Außerdem ergibt sich kein signifikantes Ergebnis für „Loudness“ oder „Breathiness“, die in der Literatur als wichtige Aspekte von Eindruckswirkung dargestellt werden.

Literatur

[1] AMON, I.: *Die Macht der Stimme*. Redline, Heidelberg, 2007.

[2] BAYERISCHES ARCHIV FÜR SPRACHSIGNALE: *PhonDat 1*, 1995.

[3] BEZOOIJEN, R. VAN: *Approximant /r/ in Dutch: Routes and feelings*. Speech Communication, 47(1):15–31, 2005.

[4] BRUCKERT, L., J. LIENARD, A. LACROIX, M. KREUTZER, and G. LEBOUCHER: *Women use voice parameter to assess men's characteristics*. Proc. Biological Sciences, 237(1582):83–89, 2006.

[5] BURKHARDT, F., R. HUBER, and A. BATLINER: *Application of speaker classification in human machine dialog systems*. In MÜLLER, C. (ed.): *Speaker Classification I: Fundamentals, Features, and Methods*, pp. 174–179. Springer, 2007.

[6] BURKHARDT, F., A. PAESCHKE, M. ROLFES, W. SENDLMEIER, and B. WEISS: *A database of German emotional speech*. In *Proc. INTERSPEECH*, pp. 1517–1520, 2005.

[7] CHATTOPADHYAY, A., D. W. DAHL, R. J. RITCHIE, and K. N. SHAHIN: *Hearing voices: The impact of announcer speech characteristics on consumer response to broadcast advertising*. Journal of Consumer Psychology, 13(3):198–204, 2003.

[8] CHILDERS, D. and A. KRISHNAMURTHY: *A critical review of electroglottography*. Proc. CRC Critical Reviews in Biomedical Engineering, 12:131–161, 1985.

[9] CUMMINGS, K. and M. CLEMENTS: *Improvements to and applications of analysis of stressed speech using glottal waveforms*. In *IEEE International Conference on Automatic Speech and Signal Processing*, pp. 25–28, 1992.

[10] D'ALESSANDRO, C.: *Voice source parameters and prosodic analysis*. In SUDHOFF, S., D. LETERNOVÀ, R. MEYER, S. PAPPERT, P. AUGURZKY, I. MLEINEK, N. RICHTER, and J. SCHLIESSER (eds.): *Language, Text and Cognition: Methods in Empirical Prosody Research*, pp. 63–87. Walter de Gruyter, Berlin, 2006.

[11] DOMMELEN, W. VAN and B. MOXNESS: *Acoustic parameters in speaker height and weight identification: Sex-specific behaviour*. Language and Speech, 38:267–287, 1995.

[12] EKLUND, R. and A. LINDSTRÖM: *Xenophones: An investigation of phone set expansion in Swedish and implications for speech recognition and speech synthesis*. Speech Communication, 35(2):81–102, 2001.

[13] FÄHRMANN, R.: *Elemente der Stimme und Sprechweise*. In: SCHERER, K. (Hrsg.): *Vokale Kommunikation*, S. 228–252. Neltz, Gießen, 1982.

[14] FOURCIN, A.: *First applications of a new laryngograph*. Medical and Biological Illustrations, 21:172–182, 1971.

[15] HENRICH, N., C. D'ALLESANDRO, M. CASTELLENGO, and B. DOVAL: *On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation*. JASA, 115(3):1321–1332, 2004.

[16] HERKNER, W.: *Lehrbuch Sozialpsychologie*. Huber, Bern, 1991.

[17] JÜRGENS, C., W. JOHANNSEN und K. FELLBAUM: *Zur Eignung von Sprechern für die Lautelemente-Bibliothek eines Sprachsynthesesystems*. In: *ITG Fachtagung Sprachkommunikation, Frankfurt/M.*, 1996.

[18] KETZMERICK, B.: *Zur auditiven und apparativen Charakterisierung von Stimmen*. Studientexte zur Sprachkommunikation. TUDpress, Dresden, 2007.

[19] MONITZ, H., A. I. MATA, I. TRANCOSO, and M. C. VIANA: *How can you use disfluencies and still sound as a good speaker?*. In *Proc. INTERSPEECH*, p. 1678, 2008.

[20] PAESCHKE, A. und W. F. SENDLMEIER: *Die Reden von Rudolf Scharping und Oskar Lafontaine auf dem Parteitag der SPD im November 1995 in Mannheim – Ein sprechwissenschaftlicher und phonetischer Vergleich von Vortragsstilen*. Zeitschrift für angewandte Linguistik, 27:5–39, 1997.

[21] SMITH, A.: *The Theory of Moral Sentiments*. Penguin Classics, 2007, 1759.

[22] STRANGERT, E. and J. GUSTAFSON: *What makes a good speaker? subjective ratings, acoustic measurements and perceptual evaluation*. In *Proc. INTERSPEECH*, pp. 1688–1691, 2008.

[23] THIEL, S.: *Die mittlere Sprechstimmlage als Wirkungsfaktor*. In: KRECH, E.-M., J. SUTTNER und E. STOCK (Hrsg.): *Ergebnisse der Sprechwirkungsforschung*, S. 202–208. Universität Halle-Wittenberg, 1987.

[24] TROUVAIN, J., S. SCHMIDT, M. SCHRÖDER, M. SCHMITZ und W. BARRY: *Modelling personality features by changing prosody in synthetic speech*. In: *Proc. Speech Prosody*, 2006.

[25] WEISS, B. and F. BURKHARDT: *Voice attributes affecting likability perception*. In *Proc. INTERSPEECH*, 2010. accepted.