Prosodische Elemente vokaler Prosodie

Benjamin Weiss

Quality and Usability Lab, TU Berlin
benjamin.weiss@tu-berlin.de

Kurzfassung: Parameter prosodischer Merkmale wurden identifiziert, die nach einschlägigen Ergebnissen mit der Valenzdimension – auch Sympathie oder auch soziale Attraktivität genannt – einhergehen. Diese Parameter wurden automatisch für je 300 männliche und weibliche Sprecher der Agender Datenbank erhoben und ausgewertet. Zwischen den Gruppen positiver, durchschnittlicher und negativer Bewertungen ergeben sich signifikante Unterschiede für Maße der Tonhöhe, Tempo und Variabilität von Tempo und Intensität. Jedoch handelt es sich bei diesen Ergebnissen nicht um ein vollständiges Set für die Valenzdimension (etwa Variabilität in Tempo, Tonhöhe oder Intensität), sondern durchaus um Parameter, die auch der Dominanzdimension zugeordnet werden. Eine strikte Trennung zwischen den Dimensionen Valenz und Dominanz für die Personenbeurteilung erweist sich demnach als unangebracht.

1 Einleitung

Der Einfluss von Stimme und Sprechweise auf interpersonelle Bewertung stellt ein breites Forschungsfeld dar. Für den Fall einer subjektiven Gesamtbewertung, also wie sympathisch oder unsympathisch ein Sprecher gehalten wird, wurden bereits Zusammenhänge mit prosodischen Merkmalen gefunden. So werden etwa tiefere Grundfrequenz und höheres Sprechtempo als positiver beurteilt [12,13k]. Da jedoch in anderen Studien erhöhte Grundfrequenz mit sympathischeren Bewertungen korrelieren, zeigt sich bereits hier die hohe Bedeutung der Beurteilungssituation, die den Interpretationsrahmen bestimmt (etwa [11]).

800 Anrufer eines telefonbasierten Sprachdialogsystems wurden anhand der mitgeschnittenen Aufnahmen von je 16 Personen auf Sympathie bewertet [4]. Sprachsignale ausgewählter extrem sympathisch und unsympathisch klingender Sprecher dieser „Agender“ Datenbank [3] wurden bereits qualitativ und quantitativ analysiert, jedoch manuell und deshalb nur für lediglich je 30 Sprecher beiden Geschlechts [12]. Hierbei zeigte sich zum einen der stärkste Unterschied zwischen den Sympathiegruppen für die dort händisch erhobene Artikulationsrate, zum anderen konnte der aufgrund der Annotation von Höreindrücken erwartete Effekt von „flacher Intonation“ parametrisch nicht bestätigt werden.

Um diese reichhaltige Datenbasis sinnvoller auszuwerten, werden nun systematisch prosodische Merkmale identifiziert, die nach einschlägigen Ergebnissen (etwa [6, 7]) die Valenzdimension interpersoneller Bewertung betreffen – also die Sympathie (vgl. Abschnitt 2). Um die oben genannten Auffälligkeiten zu überprüfen und die identifizierten Parameter systematisch zu analysieren, werden für diese Datenbank automatische Methoden verwendet, um eine weitaus größere Anzahl von Sprechern berücksichtigen zu können. Somit werden auch Parameter quantitativ analysiert, die in [14] trotz qualitativ erkannter Auffälligkeiten nicht berücksichtigt wurden (etwa Frageintonation oder Pausenanzahl).

Dieses automatische Vorgehen ermöglicht außerdem, einen nicht-linearen Zusammenhang zwischen Sympathie und Parameterausprägung zu erfassen, indem zusätzlich zu extrem (positiv

und negativ) bewerteten Sprechern auch eine Gruppe durchschnittlicher Sprecher berücksichtigt wird. Somit kann die in [14] formulierte Hypothese überprüft werden, dass Merkmale, die mit negativer Bewertung einhergehen, nicht mit solchen identisch sein müssen, die mit sympathischen Sprechern zusammenfallen.

2 Prosodische Merkmale und Bewertungsdimensionen

Prosodische Merkmale von Stimme und Sprechweise – also solche, die über die Einheiten von Phonen und Silben hinausgehen – werden für paralinguistische Fragestellungen üblicherweise in drei Gruppen unterteilt:

- *Dauern* umfassen Sprechtempo und Artikulationstempo, Pausendauern und -häufigkeit, sowie Längungs- und Kürzungsphänomene an besonderen Stellen von Äußerungen.
- *Intonation* umschreibt die Tonhöhe, also etwa Stimmlage, Stimmumfang und besondere Abfolgen, wie etwa die lokale Betonung mit der Tonhöhe, Frageintonation oder Ausmaß der Deklination.
- *Intensität* beinhaltet die allgemeine Lautstärke, die sich ohne Referenz während der Aufzeichnung später nicht auswerten lässt, und Veränderungen der Intensität zum Beispiel über die Zeit oder Differenzen zwischen verschiedenen Abschnitten oder sprachlichen Einheiten.

Ein in diesem Bereich wenig berücksichtigtes Merkmal, das durchaus auch zur Prosodie gezählt werden kann, ist der *Stimmklang* oder *Timbre*, das hier jedoch aufgrund weniger quantitativer Ergebnisse nicht berücksichtigt wird.
In einer älteren Literaturschau [6] wird trotz verwirrender Begriffsverwendung deutlich, dass die in zahlreichen Forschungsfeldern verwendeten drei Konzepte oder Dimensionen *Aktivität*, *Dominanz* und *Valenz* [9] sich auch in der sprachbasierten interpersonellen Bewertung wiederfinden lassen. So scheint Potenz mit der Zuschreibung von Kompetenz und Intensität zusammenzugehören, während Valenz oder Sympathie mit der Zuschreibung von Wohlwollen, Vertrauen und Freundlichkeit korreliere (vgl. auch [1, 8]). Die nicht berücksichtigte Aktivität erscheint hierbei intuitiv direkt für die Sprecheindrucksforschung übernehmbar zu sein, und damit identisch mit der Aufgeregtheit und ggf. Zuschreibung von Extroversion [1]. Dieses zusammenfassende Bild aus [6] entspricht hierbei durchaus dem aktuellen Stand, den [7] darstellen. Konkret zeigt sich eine positivere Bewertung (Valenz) bei

- mittlerem (oder leicht erhöhtem) Sprechtempo [12, 13] (entgegen dem positiven linearen Zusammenhang mit Dominanz),
- tieferer Sprechlage [6, 7, 15] (jedoch insbesondere Dominanz),
- höherer Variabilität der Tonhöhe [2, 10, 11],
- höhere Variabilität von Sprechtempo [10],
- höhere Variabilität der Intensität [11] und
- kürzeren Sprechpausen [6, 11] (entgegen der Anzahl bei Dominanz).

Bedeutsam ist jedoch auch ein weiterer Schluss: Das Sympathie von der vokalen Ähnlichkeit zwischen Sprecher und Bewerter abhänge [13], was jedoch im Rahmen dieses Aufsatzes nicht berücksichtigt wird.

Basierend auf diesem Stand werden für die unten beschriebenen Stimuli Parametrisierungen für *Sprechtempo, Sprechlage, Variabilität der Tonhöhe* und *Intensität*, sowie *Pausenanzahl* erhoben. Das Sprechtempo wird automatisch anhand von identifizierten Silbenkernen geschätzt, nicht anhand von Phonsegmenten, sodass bei diesen kurzen Stimuli auf die Auswertung der Variabilität verzichtet wird.

Außerdem wird aufgrund der qualitativen Bewertungen in [14] zusätzlich die Intonationskontur zur *Frageidentifikation* erhoben. Da dort auch Pausen in den kurzen Stimuli als bewusster Ausdruck eines Kommandostils interpretiert wurden, ist der Status von Pausen nicht vergleichbar mit natürlich entstehenden Strukturierungsmarkern oder ungefüllten Häsitationen innerhalb eines Sprechstils.

3 Durchführung

3.1 Verwendete Daten

Die Agender Datenbank wurde für drei Altersklassen (junge, mittlere und ältere Erwachse) und beide Geschlechter in sechs Blöcken auf Sympathie anhand einer 7-stufigen bipolaren Skala bewertet [4]. Von diesen 800 Sprechern wurden jeweils 300 pro Geschlecht ausgewählt, also die jeweils 100 mittleren, positivsten und negativsten. Die Unterscheidung nach Altersklassen wurde hierbei aufgehoben, um eine Gruppengröße zu erlangen, die den Einfluss individueller Merkmale ausschließt.

3.2 Akustische Messungen

Die Parametrisierungen der prosodischen Merkmale wurde mit Praat durchgeführt. Geschlechtsspezifische Parameter wie etwas Grundfrequenz wurden mit separaten Einstellungen erhoben. Als Parametrisierungen wurden folgende Maße ausgewählt:

- Sprechtempo: Artikulationsrate als geschätzte Silben pro Sekunde anhand von Intensitätsmaxima und Stimmhaftigkeitsinformationen [5]; auch als Differenz vom Geschlechtsmittel
- Variabilität des Sprechtempos: Die Varianz der Dauern zwischen Silbenkernen (exklusive von Pausen)
- Pausenanzahl: Die anhand der Intensität geschätzten Pausen für die Berechnung der Artikulationsrate werden direkt für die Erfassung von Pausenanzahl verwendet. Die Schwellwerte des Scripts wurden aufgrund der recht kurzen beobachteten Pausen vor betonten Wörtern verändert (minimale Pausendauer $= 150ms$, minimale Sprechdauer$= 80ms$)
- Sprechlage: Median der Grundfrequenz (in ERB)
- Variabilität der Tonhöhe: Varianz der Grundfrequenz (in ERB)
- Variabilität von Intensität: Varianz der Intensität (in dB)
- Frageintonation: Differenz der gesamten mittleren Grundfrequenz zu dem letzten 30% der Äußerung (in ERB/s)

4 Ergebnisse

Die akustischen Parameter werden separat für beide Geschlechter mit Kruskal-Wallis für die drei Gruppen der 50 positivsten, negativsten und mittleren Sympathiebewertungen ausgewertet. Alle fünf signifikanten Ergebnisse sind pro Geschlecht in Tabelle 1 abgebildet ($\alpha = .5$).

Tabelle 1 - Ergebnisse der Varianzanalysen

Geschlecht	**Parameter**	$\chi^2_{(2)}$
Frauen	Artikulationsrate	8.16*
	Varianz Rate	6.27*
	F0 Median	6.54*
Männer	F0 Median	8.56*
	Varianz Intensität	9.39**

Im Einzelnen zeigen die Ergebnisse folgende Mittelwertunterschiede, wobei signifikante Ergebnisse mit einem Verbindungsstrich markiert sind (Mann-Whitney-U Test mit α-Level-Anpassung):

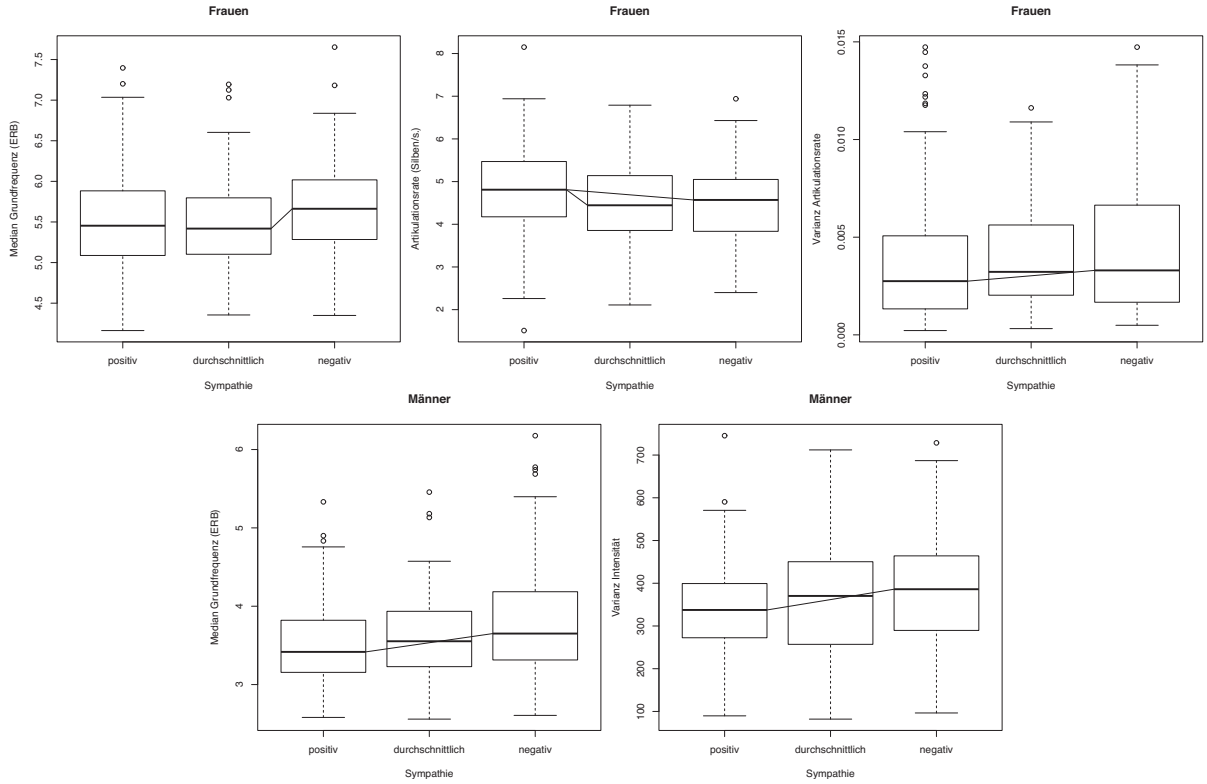

Abbildung 1 - Ergebnisse der Posthoc Tests

Kein Effekt zeigt sich für die Pausenanzahl. Da in der qualitativen Analyse in [14] die Pausen vornehmlich als Ausdruck eines Kommandostils interpretiert wurden, wird dieser Datensatz nochmals überprüft, um die Häufigkeit Stimuli mit Pausen gegenüber solchen ohne Pausen in den drei Sympathiegruppen zu überprüfen. Der χ^2 Test zeigt jedoch weder für Männer noch Frauen einen signifikanten Effekt ($\chi^2_{(2)}$= 2.3, 2.1; p=.32, .36).

Die Artikulationsrate ist für sympathische weibliche Sprecher höher als für die beiden anderen weiblichen Sprechergruppen. Eine vergleichbare Analyse mit dem Abweichen vom Mittelwert als Betrag weist für beide Geschlechter keine Signifikanz auf (p=.37 bzw. p=.21). Ebenfalls zeigt sich kein Effekt für die Parametrisierung der Frageintonation.

Jeweils vier Parameter weisen signifikante Ergebnisse für Sympathie auf. Als positiv zeigen sich demnach:

- Niedrigere Grundfrequenz für beide Geschlechter im Einklang mit der Erwartung.
- Erhöhtes Sprechtempo bei weiblichen Sprechern entgegen der Erwartung.
- Geringere Varianz in der Artikulationsrate bei weiblichen Sprechern entgegen der Erwartung.
- Geringere Varianz in der Intensität bei männlichen Sprechern entgegen der Erwartung.
- Keine Unterschiede für die Variabilität der Grundfrequenz.

5 Diskussion

Überraschenderweise zeigen sich konträre Ergebnisse für die Variabilität von Artikulationsrate (Frauen) und Intensität (Männer) im Vergleich zur Literatur. Auch das Sprechtempo, das hier bei sympathisch bewerteten Frauen höher ist, entspricht nicht der Erwartung aus der Literatur, bestätigt aber die händische Erhebung aus [14]. Lediglich die tiefere Grundfrequenz tritt erwartungsgemäß mit sympathischeren Beurteilungen für beide Geschlechter signifikant auf.

Der Abgleich mit den händisch vorgenommenen Analysen in [14] zeigt keinen Effekt für die Pausenanzahl oder die Frageintonation. Das Scheitern der quantitativen Analysen für diese beiden Beobachtungen lässt sich sicherlich auf die geringe Anzahl von Fällen zurückführen (jeweils unter 10 Fälle in [14]) und bezeugt damit das Problem, spezielle oder individuelle Ursachen für extreme Sympathiebewertungen mit korrelations- oder varianzanalytischen Verfahren nicht erfassen zu können.

Auf Basis der hier dargestellten Ergebnisse muss die Zuordnung von Wohlwollen zur Sympathie in Frage gestellt werden. Die mit wohlwollenden Sprechern auftretenden Variabilität für Tonhöhe, Intensität und Sprechtempo zeigen sich hier nicht oder sogar konträr. Beispielsweise tritt ein erhöhtes Sprechtempo und tiefere Stimmlage mit positiverer Bewertung auf, was in der Literatur eher der Dominanzdimension entspricht.

Da bei der Zuschreibung von Sympathie aufgrund geringer (akustischer) Informationen durchgeführt wird, sind Einflüsse von Stereotypen zu erwarten. Insofern scheint der hier auftretende Zusammenhang von tiefer Stimmlage und Sympathie auch für Sprecherinnen ein weiteres Mal (wie auch etwa in [15]) auf ein entsprechendes Stereotyp für Frauen hinzuweisen.

Bedeutsam ist die Asymmetrie für zwei Parameter bei weiblichen Sprechern: die mittlere Grundfrequenz ist lediglich bei Gruppe der unsympathischen Sprecherinnen erhöht, während sich für das Sprechtempo nur für die positiven Beurteilungen erhöht zeigt. Die Berücksichtigung solcher Asymmetrien sollte demnach bei Vorhersagemodellen berücksichtigt werden.

6 Fazit

Es zeigen sich signifikante Unterschiede in automatisch erhobenen Parametern der Prosodie für Gruppen von auf Sympathie bewerteten Sprechern. Allerdings entsprechen die Ergebnisse nicht der Erwartung auf Basis der Literatur, so dass die generelle Zuordnung von Sympathie als Valenzdimension zu beispielsweise Wohlwollen oder die damit einhergehenden Parameterausprägungen kritisch überprüft werden sollten.

Die hier verwendeten automatischen Verfahren wurden nicht validiert, zeigen aber am Beispiel der Artikulationsrate vergleichbare Ergebnisse mit – für eine Untermenge händisch erhobenen – Werte. Die hier aufgeführten Ergebnisse und auch die zugrundeliegende automatische Methode

zur Pausen- und Sprechtemposchätzung sollen in Kürze anhand eines bereits händisch annotierten Korpus validiert werden.

7 Danksagung

Diese Arbeit wurden von der DFG gefördert (WE 5050/1-1).

Literatur

[1] BURGOON, J. K.: *Attributes of the newscaster's voice as predictors of his credibility*. Journalism Quarterly, 55:276–281, 1978.

[2] BURGOON, J. K. und L. AHO: *Three field experiments on the effects of violations of conversational distance*. Communication Monographs, 49:71–88, 1982.

[3] BURKHARDT, F., M. ECKERT, W. JOHANNSEN und J. STEGMANN: *A Database of Age and Gender Annotated Telephone Speech*. In: *Proc. LREC*, 2010.

[4] BURKHARDT, F., B. SCHULLER, B. WEISS und F. WENINGER: *Would You Buy A Car From Me? On the Likability of Telephone Voices*. In: *Proc. Interspeech*, 2011.

[5] JONG, N. H. DE und T. WEMPE: *Praat script to detect syllable nuclei and measure speech rate automatically*. Behavior Research Methods, 41:385–390, 2009.

[6] KETROW, S. M.: *Attributes of a telemarketer's voice and persuasiveness: A review and synthesis of the literature*. Journal of Direct Marketing, 4:7–21, 1990.

[7] KREIMAN, J. und D. VAN LANCKER SIDTIS: *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception*. Wiley, Chichester, 2011.

[8] MEHRABIAN, A.: *Some referents and measures of nonverbal behavior*. Behavioral Research Methods and Instrumentation, 1:213–217, 1969.

[9] OSGOOD, C., G. SUCI und P. TANNENBAUM: *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception*. Univ. of Illinois Press, 1957.

[10] RAY, G. B.: *Vocally cued personality proto-types: An implicit personality theory approach*. Communication Monographs, 53:266–276, 1986.

[11] SCHERER, K. R.: *Personality Markers in Speech*. In: SCHERER, K. R. und H. GILES (Hrsg.): *Social Markers in Speech*, S. 147–209. Cambridge University Press, 1980.

[12] SMITH, B. L., B. L. BROWN, W. J. STRONG und A. C. RENCHER: *Effects of speech rate on personality perception*. Language and Speech, 18:145–253, 1975.

[13] STREET, R. L. und R. M. BRADY: *Speech rate acceptance ranges as a function of evaluative domain, listener speech rate, and communicative Context*. Communication Monographs, 49:290–308, 1982.

[14] WEISS, B. und F. BURKHARDT: *Is 'not bad' good enough? Aspects of unknown voices' likability*. In: *Proc. Interspeech*, 2012.

[15] WEISS, B. und S. MÖLLER: *Wahrnehmungsdimensionen von Stimme und Sprechweise*. In: *Proc. Elektronische Sprachsignalverarbeitung (ESSV)*, S. 261–268, 2011.