INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Social Attractiveness in Dialogs

Antje Schweitzer, Natalie Lewandowski, Daniel Duran

Institute for Natural Language Processing, University of Stuttgart, Germany
⟨firstname.lastname⟩@ims.uni-stuttgart.de

Abstract

This study investigates how acoustic and lexical properties of spontaneous speech in dialogs affect perceived social attractiveness in terms of speaker likeability, friendliness, competence, and self-confidence. We analyze a database of longer spontaneous dialogs between German female speakers and the mutual ratings that dialog partners assigned to one another after every conversation. Thus the ratings reflect long-term impressions based on dialog behavior. Using linear mixed models, we investigate both classical acoustic-prosodic and lexical parameters as well as parameters that capture the degree of speakers' adaptation, or "convergence", of these parameters to each other. Specifically we find that likeability is correlated with the speaker's lexical convergence as well as with her convergence in f_0 peak height. Friendliness is significantly related to variation in intensity. For competence, the proportion of positive words in the dialog, variation in shimmer, and overall phonetic convergence are significant correlates. Self-confidence finally is related to several prosodic, phonetic, and lexical adaptation parameters. In some cases, the effect depends on whether interlocutors also had eye contact during their conversation. Taken together, these findings provide evidence that in addition to classical parameters, convergence parameters play an important role in the mutual perception of social attractiveness.

Index Terms: social attractiveness, convergence, spontaneous speech

1. Introduction

A considerable body of literature has investigated acoustic correlates of extralinguistic factors such as emotion and personality (see e.g. [1, 2] for overviews). In recent years, voice attractiveness and pleasantness have also gained interest, and they have been hypothesized to depend on similar parameters [3]. In this field the focus is often on cross-gender perception of voice attractiveness, and the notion of attractiveness in these cases is then to some extent biased to sexual attractiveness (see for instance [4, 5, 6, 7, 8]). Other research on voice attractiveness, or pleasantness, is in the context of dialog systems aiming to provide pleasant synthetic voices [9, 10]. In any case, research in the domain of vocal attractiveness or pleasantness typically makes use of short, often read, sometimes synthetic, stimuli, which are rated by independent listeners outside communication situations (e.g. [5, 11, 12, 13, 10, 6, 14, 8, 15]); [3] refer to this as a "passive rating scenario" . One of the very few exceptions is [7] who investigate mutual ratings of participants of a speed dating game.

Among the parameters that have been suggested to be related to voice attractiveness and pleasantness, many are related to f_0; for instance [4] finds closely spaced low-frequency harmonics correlate with women's perception of male attractiveness, [5] show that feminine women prefer lower-pitched male voices; in a speed-dating study [7] women perceived as friendly exhibit higher maximum pitch and greater pitch variance, [8] find that men prefer higher pitch in women, while women prefer lower pitch in men; [11] looking at male voices also find that low pitch is preferred, and additionally report an effect of f_0 variance in interaction with f_0 mean, where low variance is preferred for voices with medium mean f_0, while mid and high variance is preferred for either low or high mean f_0 voices.

The finding that f_0 or pitch play an important role is challenged by [15] who find that low F1 is rated as more attractive for male voices, while there is no significant effect of f_0. They claim that voices that are thought to reflect greater body height in men are preferred, and since low F1 as an indicator of vocal tract size is a more reliable estimator of body height than f_0, this explains the null result regarding f_0. In addition [15] find breathier female voices, and male voices with shorter durations to be more attractive. They also report a possibly sociophonetic effect, namely the preference of a low F2 in /u/ for female voices, which they interpret as an indication that the /u/ fronting that is typical for female Californian speakers is evaluated positively by raters.

Other studies corroborate the relevance of vocal tract length as derived from formants, for instance [5] find a preference for more dispersed formants in female voices, and [13] a preference for less dispersed formants in male voices when evaluated by females; however the latter also find that dispersion is not predictive of dominance ratings of male voices by men. The preference for breathier female voices found by [15] had previously been found by [8], interestingly along with a preference for wide formant dispersion in female voices, contradicting [13] and in contrast to [15] who obtained this finding for men but not for women. A few studies report results for intensity, e.g. more varied intensity in men who are perceived as friendly [7]. That particular study also finds effects of lexical parameters and laughter, for instance friendly men and women are reported to use more turn-medial or turn-final laughter. Another study [16] reports an effect of speech rate (higher likeability scores for decreased speech rate), but only for women. That study also finds that skewness correlates with likeability for women, while the "speaker's formant" (increased intensity in long-term average spectrum between 3 and 4 kHz) is found in the voices of likeable men.

Another line of research is interested in the question of whether it is possible, and with which accuracy, to tell if voices will be perceived as attractive or pleasant, given a very large set of (usually acoustic-phonetic) parameters [10, 9, 3]. These studies typically investigate many more parameters; for instance, [10] use more than 4000 features derived from 60 low-level cepstral, auditory spectral, energy-related, voice-related, or f_0-related parameters extracted with openSMILE [17]; [9] use 310 features comprising energy-related features as well as f_0 formants, cepstral features, voice quality, and articulation speed. However while these studies to some extent also investigate which features are central for the prediction, they do usually not

http://dx.doi.org/10.21437/Interspeech.2017-833

aim at isolating the effects of specific parameters on perceived attractiveness.

The present study complements existing research in several ways: (i) the attractiveness ratings in our case are ratings that conversation partners gave each other after a longer conversation, and they are in this respect more "authentic" than ratings collected in a passive rating scenario from independent raters; (ii) since the ratings are obtained on the basis of a longer dialog instead of a relatively short excerpt of speech they may not only reflect relatively superficial acoustic properties but also more long-term parameters including conversation content; and (iii) we expect that in addition to more superficial, physically-driven acoustic parameters, conversation "behavior" including adjustments of a speaker to the partner, may affect (or reflect) the partner's rating to some extent. Since our ratings are based on "more evidence" than the classic ratings from passive rating scenarios (in that they are based on mutual perception in longer dialogs including behavior in the dialog), we suggest that they reflect "social attractiveness" rather than the classic "voice attractiveness". It should also be noted that we only look at female-female conversations, thus our ratings are less likely to reflect sexual aspects of attractiveness.

We describe the data we used in this study in the following section, followed by the results of our statistical analysis in section 3 and their discussion in section 4.

2. Method

We use the GECO database of German conversational speech [18, 19] to investigate segmental, prosodic, and lexical correlates of perceived social attractiveness in interactions between women. This database was originally recorded to investigate phonetic convergence, i.e. to investigate how speakers in natural, free conversations adapt their way of speaking to each other. In addition to the conversations themselves, GECO also contains mutual ratings of the speakers, including aspects such as likeability and friendliness. We use these data to explore the relation between these ratings and the phonetic and lexical parameters that the speakers employed. We will describe the speech data, the ratings, and the parameters we extracted in the following subsections.

2.1. Speech Data

GECO contains 46 dialogs between 13 previously unacquainted German female speakers in different dialog pairings. Each dialog lasted around 25 minutes, and the speech data amounts to approx. 21 hrs of speech. The dialogs were completely free, spontaneous dialogs about topics of the speakers' choice, recorded in a sound-attenuated chamber in two conditions; a *monomodal* condition where participants were separated by a sound-treated wall and could not see each other, and a *multimodal* condition where participants were only separated by a transparent screen and could see each other.

2.2. Ratings

After each dialog, speakers filled in questionnaires about their impression of the conversation partner and of the conversation in general. In this study, we use items related to perceived social attractiveness, namely ratings of how *likeable* ("sympathisch"), *friendly* ("freundlich"), *competent* ("kompetent"), and *self-confident* ("selbstsicher") speakers were perceived by their partner. Answers to each of these items were on a 5-point Likert scale ranging from "yes" to "no".

2.3. Extracted parameters

2.3.1. Segmental and prosodic parameters

In order to receive reliable measurements of the phonetic parameters, especially the voice quality parameters, we restricted our analyses to syllables with quantitatively long vowels (i.e. vowels with a /:/ symbol) which were at least 0.1 s long. We extracted properties related to pitch using two different approaches: Once we measured the pitch median in each long vowel, and once we used the PaIntE f_0 parametrization method [20, 21] to extract the amplitudes of f_0 rises and falls as well as the height of f_0 peaks in the corresponding syllable. As a measure of temporal prosodic properties, we extracted syllable durations. Regarding voice quality, we used Praat [22] to calculate shimmer, jitter, and the harmonics-to-noise ratio (HNR) in each long vowel using the voice report function. We also measured intensity in each long vowel. In addition we calculated F1 and F2 using *formant* from the ESPS/waves package. Shimmer, jitter, HNR and intensity were scaled and centered on a by-vowel basis to avoid vowel-intrinsic effects on each parameter. We then eliminated outliers following standard procedure by discarding all data points where for at least one dimension the value was more than 1.5 times the Interquartile Range beyond the quartiles, reducing the number of data points from 23,244 to 11,332. In case of the pitch parameters, we did this on a by-speaker basis; in case of the voice quality and formant parameters, on a by-speaker, by-vowel basis. We then calculated means and standard deviations of all these parameters for each speaker and each conversation. We refer to the resulting mean parameters below as *pitch*, *f0rise*, *f0fall*, *f0peak*, *syldur*, *shimmer*, *jitter*, *HNR*, and *intensity* and to their standard deviations as *pitch.sd*, *f0rise.sd*, *f0fall.sd*, *f0peak.sd*, *syldur.sd*, *shimmer.sd*, *jitter.sd*, *HNR.sd*, and *intensity.sd*. The means and standard deviations of the formants were only used for calculating the convergence parameters below.

2.3.2. Phonetic convergence parameters

In addition to the above parameters, we wanted to capture how speakers adapted these parameters to their partners, since it is well accepted that the degree of adaptation (and its direction, i.e., divergence or convergence) is related to social factors (e.g. [23, 24, 25]). For the intonation and formant parameters above, we calculated the adaptation as the difference between the speaker's mean in a specific dialog and the speaker's mean in all her other dialogs, normalized by her standard deviation in her other dialogs. We multiplied by -1 in cases where the partner's mean was below the speaker's mean to ensure that positive values occur in cases where the speaker adapted her mean towards the other speaker's mean (i.e. convergence), and negative values in cases where the speaker adapted away from the other speaker (i.e. divergence). We refer to the resulting convergence parameters below as *f1.conv*, *f2.conv*, *pitch.conv*, *f0rise.conv*, *f0fall.conv*, *f0peak.conv*, and *syldur.conv*. Since speakers may employ individual parameters in convergence, we added two more features, which correspond to the highest and lowest convergence value observed for a speaker in a dialog. The highest value thus captures the strongest convergence a speaker exhibited in any of the parameters (*conv*), and the lowest captures the strongest divergence exhibited (*div*).

2.3.3. Lexical convergence and similarity

We defined a measure for lexical convergence which is based on the well-known *tf-idf* vector-space measure for document similarity [26]. The tf-idf measure captures how central a specific word is to a conversation, compared to other conversations. We treated all *half-dialogs* (all tokens in a dialog uttered by the same speaker) as individual documents. We then determined tf-idf values for all normalized (all lower-case) tokens in these half-dialogs using the *tidyverse* package [27] in R. Each half-dialog is then represented by a vector where the dimensions of the vector correspond to the word types of all dialogs in the corpus. The values of the vectors correspond to the tf-idf values, with 0 for all tokens which do not appear in that half-dialog.

For each half-dialog we computed an *overlap score* [26], with the word vector of the half-dialog of the speaker corresponding to the *query* and the vector representing the half-dialog of the listener corresponding to the *document*. Thus we get two such overlap scores for each dialog, one for each speaker, called *overlap* below. We also computed similarity of the word vectors using the Euclidean distance: For each dialog, we computed distances between the vectors corresponding to the half-dialogs of the two speakers. Smaller distances indicate higher similarity between the words the speakers use within the dialog. For each speaker we also computed the tf-idf vector of the words spoken by this speaker in her other dialogs. This served as a reference for normalizing the distance between the two speakers relative to the speakers' "self-distances". From these distances, we calculated three features: the absolute distance between the two speakers, *dist*, their distance normalized by the distance between their other dialogs, *dist.rel*, and their distance normalized by the respective speaker's self-distance, *selfdist*.

2.3.4. Lexical content and sentiment

As an approximate sentiment analysis of the spoken words, we assigned positive and negative polarity to all tokens in the dialog transcripts using data from [28, 29]. We did not explicitly take into account neutral sentiment polarity. We computed the ratio of *positive* and *negative* words for each half-dialog. In addition, we looked at the absolute number and the proportion of tokens that each speaker contributed to the dialog (*ntok* and *ntok.rel*) as well as at occurrences of laughter both in absolute and in relative terms (*laugh*, *laugh.rel*).

3. Results

The analysis was conducted using R version 3.3.2 [30] and the packages *lme4* [31] and *afex* [32]. Data manipulation and exploration was performed with the *dplyr* package [33] from *tidyverse* [27]. All continuous variables were z-scored and all categorical variables centered. The best random structure for each dependent variable was determined first, by running multiple anova model comparisons.

We then fitted full models, containing all acoustic, lexical and sentiment variables, for the four social scores *self-confident, competent, likeable, and friendly* as the dependent variables. The full models were then reduced stepwise by eliminating the parameter with the highest variance inflation factor (vif) in each new calculation turn [34] until all vifs$\leq$4. We then fit models with the remaining factors, once with, and once without an interaction with modality (*mod*) and compared with an anova. Finally, the better model of the two was further reduced by considering model fit. The impact of every eliminated variable on the linear mixed model was tested with an anova model comparison – in case it did not improve the model significantly, it was removed, until the best possible model fit was reached (i.e. that with significantly lower Akaike information criterion *AIC*). Tables 1 through 4 list model terms and estimates of the coefficients for fixed effects along with p-values obtained with the *mixed* function from the *afex* package [32]. Significant factors are highlighted in gray.

Table 1: *Fixed effects in the self-confidence model.*

Parameter	Estimate	SE	t-value	p
f0rise.conv	0.1862	0.0842	2.21	.04*
f1.conv	-0.0654	0.0858	-0.76	.45
syldur.conv	0.2497	0.0904	2.76	.007*
dist	-0.1711	0.0821	-2.08	.04*
overlap	0.2388	0.0981	2.43	.02*
mod(multi)	0.5698	0.1313	4.34	<.0001*
f0rise.conv:mod	-0.4525	0.1276	-3.55	.0007*
f1.conv:mod	0.3664	0.1205	3.04	.003*
syldur.conv:mod	-0.2631	0.1299	-2.03	.05
dist.rel:mod	0.0961	0.1491	0.65	.54
overlap:mod	-0.2557	0.1297	-1.97	.05

self-confidence $\sim$ *(f0rise.conv + f1.conv + syldur.conv + dist.rel + overlap) * mod + (1|partner)*

Table 2: *Fixed effects in the competence model.*

Parameter	Estimate	SE	t-value	p
positive	0.2914	0.1022	2.85	.006*
laugh.rel	-0.2075	0.1004	-2.07	.04*
syldur.sd	0.2559	0.1282	1.99	.06*
shimmer.sd	-0.2166	0.0960	-2.26	.03*
conv	0.1715	0.0693	2.47	.02*
mod(multi)	0.5125	0.1255	4.08	.0001*
positive:mod	-0.4679	0.1309	-3.58	.0008*
laugh.rel:mod	1.6518	0.8074	2.05	.04*
syldur.sd:mod	-0.3356	0.1635	-2.05	.05
shimmer.sd:mod	0.1443	0.1543	0.94	.36
conv:mod	-0.0779	0.1094	-0.71	.48

competence $\sim$ *(positive + laugh.rel + syldur.sd + shimmer.sd + conv) * mod + (1|partner) + (1|speaker)*

As can be seen in Table 1, speakers who are rated as more *self-confident* converge more with respect to f_0 rise amplitudes (*f0rise.conv*) and with respect to syllable durations (*syldur.conv*). They exhibit smaller distance to their partner in terms of their word vectors (*dist*) and a greater overlap in terms of tf.idf vectors (*overlap*). We also observe a main effect of modality, i.e. speakers are rated as more self-confident in general when they could see each other during the conversation. There are also two-way interactions of *f0rise.conv* and *f1.conv* with modality showing that the convergence effect with respect to f_0 rise amplitudes is cancelled out in the multimodal condition, and that instead, we observe convergence in terms of F1 (*f1.conv*) in the multimodal condition.

With respect to *competence* (Table 2) we observe that convergence behavior seems to be less important, as only one significant factor is related to convergence: speakers who exhibit stronger convergence behavior in at least one phonetic dimension (*conv*) are rated as more competent. In addition, a number of factors not related to convergence play a role: Speakers who use more positive words, but laugh less, are rated as more competent in monomodal dialogs. Again, both effects are in the

Table 3: *Fixed effects in the friendliness model.*

Parameter	Estimate	SE	t-value	p
intensity.sd	-0.1103	0.0456	-2.42	.02*
friendliness $\sim$ intensity.sd + (1|partner)				

Table 4: *Fixed effects in the likeability model.*

Parameter	Estimate	SE	t-value	p
f0peak.conv	0.1271	0.0936	1.36	.18
f2.conv	0.1239	0.0850	1.46	.15
f1.conv	0.0368	0.0849	0.43	.67
intensity.sd	-0.1391	0.0889	-1.56	.12
dist.rel	-0.1923	0.0874	-2.20	.03*
mod(multi)	0.0717	0.1387	0.52	.61
f0peak.conv:mod	-0.3080	0.1257	-2.45	.02*
f2.conv:mod	-0.1304	0.1228	-1.06	.29
f1.conv:mod	0.1364	0.1171	1.17	.25
intensity.sd:mod	-0.0177	0.1234	-0.14	.89
dist.rel:mod	0.2179	0.1493	1.46	.15
likeability $\sim$ *(f0peak.conv + f2.conv + f1.conv + intensity.sd + dist.rel) * mod + (1|partner)*				

opposite direction in the multimodal condition, as can be seen from the negative coefficients in the interaction of both factors with modality (*mod*). In addition, speakers with less shimmer are rated as more competent. As before, there is also a main effect of modality indicating that speakers are rated more competent in general in the multimodal condition.

The model for *friendliness* (see Table 3) indicates that only one factor is significantly related to friendliness: speakers who vary their intensity less are rated as more friendly. Finally, we find that two factors have a significant effect on perceived *likeability* (Table 4), and they are again related to convergence behavior: Speakers who have a smaller relative tf-idf distance (*dist.rel*) to their partners, i.e. converge more with respect to words that are central, are perceived as more likeable. Interestingly, convergence in terms of adapting f_0 peak height is also relevant, although not in general, but in the multimodal condition, as can be seen from the interaction between *f0peak.conv* and *mod*. The direction of the effect is also unexpected: we observe that speakers who adjust their f_0 peak height less to their partners are rated as more likeable.

4. Discussion and Conclusion

In general, we find that a number of the proposed parameters are related to social attractiveness as captured by the four variables *self-confidence*, *competence*, *friendliness*, and *likeability*. Among these are some well-known parameters, for instance we confirm an effect of variability in intensity on friendliness [7]. That study also observed lower use of negative emotion and more turn-medial or turn-final laughter for friendly speakers, which is similar to our finding that more positive words and more laughter are received more positively, even though in our case these parameters are associated with higher competence rather than friendliness. It has been argued that likeability may be related both to benevolence and to status [16] so the fact that friendliness and self-confidence correlate with similar parameters in [7] and in our study may be further evidence that the two concepts cannot be clearly separated.

In any case, most parameters that have been found to correlate with aspects of voice attractiveness in earlier studies do not significantly contribute to social attractiveness in our data. Specifically, we find no effect of absolute pitch or pitch range (which would be captured by the standard deviation of the pitch-related parameters). We also do not find effects of the traditional voice parameters, except that variation in the amount of shimmer is negatively related to competence.

Instead, most factors that we find to be significant in predicting social attractiveness are convergence factors. Ignoring modality and its interactions with the other fixed factors, 6 out of 11 significant factors are related to either phonetic or lexical convergence. In case of self-confidence and likeability, they are clearly the dominant factors. This is not true for the competence model, but even there the factor *conv*, i.e. the value of the strongest convergence observed among the phonetic parameters, is significant. The friendliness model is the exception, however this models stands out anyway since it is surprising that friendliness is best modeled using only one parameter (variation in intensity). Of the convergence parameters, the ones related to intonation and rhythm (f_0 parameters and syllable durations) seem to be most common.

Looking at the implications of the models presented above, it is interesting to note that self-confidence is related to convergence in several ways. After all, it could have been expected that self-confident speakers have the least motivation to converge to their partners. A further interesting finding is that competence is, among other factors, related to the use of more positive tokens in the dialog—simplified somewhat, this could imply that we perceive more optimistic people as more competent. And finally, likeability is consistently related to lexical convergence, and only in an interaction (negatively) related to phonetic convergence. This could be interpreted as indicating that for the perception of likeability, content is more important than its phonetic-prosodic realization. Before concluding, it should be noted that three of the four models indicate an effect of modality on the attractiveness scores, and this effect is positive in all cases. This may be due to the fact that it is easier to connect to a dialog partner if there is eye contact, and this could be reflected in the overall higher ratings in the multimodal condition.

To conclude, this paper investigates a less well-researched aspect of attractiveness, namely “authentic” social attractiveness as perceived in real, natural interactions. We show that convergence seems to be an important factor in speakers' mutual perception, maybe more important than the phonetic and prosodic parameters traditionally investigated in studies on attractiveness. The present work investigates only female speakers, which implies that the ratings are probably relatively unaffected by sexual attractiveness. However, it precludes generalization to male speakers. This can be addressed in the future, as we are currently annotating the GECO 2 database, which will contain conversations in the same setup as in the present GECO database, but will involve both mixed-gender and single-gender conversations [35].

5. Acknowledgements

This work was funded by the German Research Foundation (DFG) within the Collaborative Research Center SFB 732 / A4.

6. References

[1] K. Scherer, "Vocal communication of emotion: A review of research paradigms," *Speech Communication*, vol. 40, no. 1-2, pp. 227–256, Apr. 2003.

[2] J. Kreiman and D. Sidtis, *Foundations of voice studies: an interdisciplinary approach to voice production and perception*. Chichester: Wiley-Blackwell, 2011.

[3] B. Schuller, S. Steidl, A. Batliner, E. Nöth, A. Vinciarelli, F. Burkhardt, R. van Son, F. Weninger, F. Eyben, T. Bocklet, G. Mohammadi, and B. Weiss, "A survey on perceived speaker traits: Personality, likability, pathology, and the first challenge," *Computers Speech & Language*, 2014.

[4] S. A. Collins, "Men's voices and women's choices," *Animal Behaviour*, vol. 60, pp. 773–780, 2000.

[5] D. Feinberg, B. Jones, M. L. Smith, F. Moore, L. DeBruine, R. Cornwell, S. Hillier, and D. Perrett, "Menstrual cycle, trait estrogen level, and masculinity preferences in the human voice," *Hormones and Behavior*, vol. 49, no. 2, pp. 215 – 222, 2006.

[6] D. A. Puts, D. H. Bailey, R. A. Cárdenas, R. P. Burriss, L. L. Welling, J. R. Wheatley, and K. Dawood, "Women's attractiveness changes with estradiol and progesterone across the ovulatory cycle," *Hormones and Behavior*, vol. 63, no. 1, pp. 13 – 19, 2013.

[7] R. Ranganath, D. Jurafsky, and D. A. McFarland, "Detecting friendly, flirtatious, awkward, and assertive speech in speed-dates," *Computer Speech and Language*, vol. 27, pp. 89–115, 2013.

[8] Y. Xu, A. Lee, W.-L. Wu, and P. B. Xuan Liu, "Human vocal attractiveness as signaled by body size projection," *PLoS ONE*, vol. 8, no. 4, 2013.

[9] L. Coelho, D. Braga, and C. García-Mateo, "An automatic voice pleasantness classification system based on prosodic and acoustic patterns of voice preference," in *Proceedings of the 12th Annual Conference of the International Speech Communication Association (Interspeech, Florence, Italy)*, 2011.

[10] F. Burkhardt, B. Schuller, B. Weiss, and F. Weninger, "Would you buy a car from me? – on the likeability fo telephoe voices," in *Proceedings of the 12th Annual Conference of the International Speech Communication Association (Interspeech, Florence, Italy)*, 2011, pp. 1557–1560.

[11] D. Riding, D. Lonsdale, and B. Brown, "The effects of average fundamental frequency and variance of fundamental frequency on male vocal attractiveness to women," *Journal of Nonverbal Behavior*, vol. 30, no. 2, pp. 55–61, 2006.

[12] L. Bruckert, P. Bestelmeyer, M. Latinus, J. Rouger, I. Charest, G. A. Rousselet, H. Kawahara, and P. Belin, "Vocal attractiveness increases by averaging," *Current Biology*, vol. 20, no. 2, pp. 116 – 120, 2010.

[13] C. R. Hodges-Simeon, S. J. C. Gaulin, and D. A. Puts, "Different vocal parameters predict perceptions of dominance and attractiveness," *Human Nature*, vol. 21, no. 4, pp. 406–427, 2010.

[14] B. Weiss, "Prosodische Elemente vokaler Prosodie," in *Elektronische Sprachsignalverarbeitung 2013 – Tagungsband der 24. Konferenz*, P. Wagner, Ed. Dresden: TUDpress, 2013, pp. 212–217.

[15] M. Babel, G. McGuire, and J. King, "Towards a more nuanced view of vocal attractiveness," *PLoS ONE*, vol. 9, no. 2, 2014.

[16] B. Weiss, "Akustische Korrelate von Sympathieurteilen bei Hörern gleichen Geschlechts," in *Elektronische Sprachsignalverarbeitung 2015: Tagungsband der 26. Konferenz, Eichsttt, 25. - 27. Mrz 2015*, G. Wirsching, Ed. Dresden: TUDpress, 2015, pp. 165–171.

[17] F. Eyben and B. S. Martin Wöllmer, "openSMILE – the munich versatile and fast open-source audio feature extractor," in *Proc. ACM Multimedia (MM), ACM*, 2010.

[18] A. Schweitzer and N. Lewandowski, "Convergence of articulation rate in spontaneous speech," in *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013, Lyon)*, 2013, pp. 525–529.

[19] ——, "Social factors in convergence of f1 and f2 in spontaneous speech," in *Proceedings of the 10th International Seminar on Speech Production, Cologne*, 2014.

[20] G. Möhler and A. Conkie, "Parametric modeling of intonation using vector quantization," in *Proceedings of the Third International Workshop on Speech Synthesis (Jenolan Caves, Australia)*, 1998, pp. 311–316.

[21] G. Möhler, "Improvements of the PaIntE model for F0 parametrization," Manuscript, 2001.

[22] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," Mar. 2017, version 6.0.24. [Online]. Available: http://www.praat.org/

[23] H. Giles and P. M. Smith, "Accommodation theory: Optimal levels of convergence," in *Language and Social Psychology*, H. Giles and R. St. Clair, Eds. Oxford: Blackwell, 1979, pp. 45–65.

[24] R. Street, "Speech convergence and speech evaluation in fact-finding interviews," *Human Communication Research*, vol. 11, no. 2, pp. 139–169, 1984.

[25] J. S. Pardo, R. Gibbons, A. Suppes, and R. M. Krauss, "Phonetic convergence in college roommates," *Journal of Phonetics*, vol. 40, no. 1, pp. 190–197, 2012.

[26] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to information retrieval*. New York: Cambridge University Press, 2008.

[27] H. Wickham, *tidyverse: Easily Install and Load 'Tidyverse' Packages*, 2017, R package version 1.1.1. [Online]. Available: https://CRAN.R-project.org/package=tidyverse

[28] U. Waltinger, "Germanpolarityclues: A lexical resource for german sentiment analysis," in *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC)*. Valletta, Malta: electronic proceedings, May 2010.

[29] ——, "Sentiment analysis reloaded: A comparative study on sentiment polarity identification combining machine learning and subjectivity features," in *Proceedings of the 6th International Conference on Web Information Systems and Technologies (WEBIST '10)*, Valencia, Spain, April 2010.

[30] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2016. [Online]. Available: https://www.R-project.org/

[31] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[32] H. Singmann, B. Bolker, J. Westfall, and F. Aust, *afex: Analysis of Factorial Experiments*, 2016, r package version 0.16-1. [Online]. Available: https://CRAN.R-project.org/package=afex

[33] H. Wickham and R. Francois, *dplyr: A Grammar of Data Manipulation*, 2016, r package version 0.5.0. [Online]. Available: https://CRAN.R-project.org/package=dplyr

[34] F. F. Austin, "Vif.mer function," 2014. [Online]. Available: https://raw.githubusercontent.com/aufrank/R-hacks/master/mer-utils.R

[35] A. Schweitzer, N. Lewandowski, D. Duran, and G. Dogil, "Attention, please! – Expanding the GECO database," in *Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow*, 2015, paper number 620.