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

Towards Speaker Characterization: Identifying and Predicting Dimensions of Person Attribution

Laura Fernández Gallardo, Benjamin Weiss

Quality and Usability Lab, Telekom Innovation Laboratories, TU Berlin, Germany
(laura.fernandezgallardo|benjamin.weiss)@tu-berlin.de

Abstract

A great number of investigations on person characterization rely on the assessment of the Big-Five personality traits, a prevalent and widely accepted model with strong psychological foundation. However, in the context on characterizing unfamiliar individuals from their voices only, it may be hard for assessors to determine the Big-Five traits based on their first impression. In this study, a 28-item semantic differential rating scale has been completed by a total of 33 listeners who were presented with 15 male voice stimuli. A factor analysis on their responses enabled us to identify five perceptual factors of person attribution: (social and physical) attractiveness, confidence, apathy, serenity, and incompetence. A discussion on the relations of these dimensions of speaker attribution to the Big-Five factors is provided and speech features relevant to the automatic prediction of our dimensions are analyzed, together with SVM regression performance. Although more data are needed to validate our findings, we believe that our approach can lead to establish a space of person attributions with dimensions that can easily be detected from utterances in zero-acquaintance scenarios.

Index Terms: speaker characterization, semantic differential, automatic prediction

1. Introduction

The characterization of individuals from their voice has attracted the attention of researchers in the last few decades. Whereas other voice technologies are concerned with the conveyed message (speech recognition and speech synthesis), it is also possible to detect talkers' attributes such as age, gender, weight, height, and social characteristics such as personality and likability. The automatic classification of speakers is of interest for adaptive human-machine dialogs [1], whereas this research can also benefit speech synthesis based on acoustics revealing human traits, e.g. for conversational agents [2]. Moreover, understanding and predicting human development, behavior, and group interactions may also have their foundation on speaker characterization techniques.

The Brunswikian Lens model is commonly applied to the processes of expression and perception of speaker traits such as extroversion [3]. According to this model, research may focus on the inference of self-assessments or on the inference of assessments made by listeners. In this paper, we address the second approach. Numerous previous works have also concentrated on finding acoustic cues that influence listeners' attribution of speaker characteristics to unfamiliar persons. The ultimate goal of these investigations is to be able to predict those traits automatically, bridging the gap between low-level information (e.g. speech features) and high-level information of individuals' perceived characteristics.

The Big-Five personality paradigm is very frequently adopted to characterize speakers [4]. Well-known corpora such as the SSPNet Speaker Personality Corpus (SPC) and the Electronically Activated Recorded (EAR) corpus have been labeled by external raters employing questionnaires to gather Big-Five attributions, such as the Big-Five inventory of 10 items (BFI-10) [5].

While the Big-Five traits are well-established to describe personality, human labelers might have difficulties to rate the different items of Big-Five questionnaires when confronted with unknown voice stimuli. It has already been observed that, in contrast to *Extraversion* and *Neuroticism*, less stereotyped personality traits such as *Openness to Experience* are harder to perceive (and to later automatically predict) from speech [6]. The personality classification and regression techniques examined in [7] performed best for the *Extraversion* trait reported by observers (unfamiliar with the speakers) while the models for *Openness to Experience* did not outperform the baseline. Hence, self-reports, personality assessments based on audio-visual material, or personality assessed by the target' close friends or family [8, 9] can provide more reliable scores when using the Big-Five model.

Another widely used model for representing interpersonal dispositions is the interpersonal circumplex, with the two orthogonal dimensions of *competence/dominance* and *warmth/benevolence*. These major personality factors can be derived by e.g. the use of adjective scales [10]. The investigations in [11, 12] examine personality factors that were obtained from listeners' judgments on semantic differential rating scales. The two factors derived these works match *competence* and *benevolence* of the interpersonal circumplex. Our work follows a similar approach, yet incorporating an extended set of adjectives to detect persons' attributions from their voices. As will be described in Section 2, our adjective scales attempt to cover additional voice personality aspects than those represented by the interpersonal circumplex. Accordingly, more than two factors have been detected in our analysis (Section 3).

Different to personality attributions, other studies have examined speaker social characteristics such as charisma [13], persuasion [14], speakers' expertise [15], and voice likability [16]. A common motivation across these works was to detect factors or speech features influencing listeners' impressions.

In this paper, we attempt to find dimensions of speaker attributions based on a 28-item semantic differential questionnaire and subsequent factor analysis, considering the voices of 15 male German speakers. The perceptual dimensions found are then related to externally attributed Big-Five personality. Besides, speech features contributing to the prediction of the identified perceptual dimensions are analyzed. The final goal of this research is to be able to measure and represent subjective person attributions measured from observers' first impressions based on speech only, as an alternative to the more general Big-Five traits.

http://dx.doi.org/10.21437/Interspeech.2017-328

2. Questionnaire and listening test

We employed a questionnaire involving a 28-item semantic differential rating scale. The attribute opposition pairs were presented in German at the ends of continuous sliders, which enables simple judgments from voice samples.

This questionnaire aims at assessing major interpersonal attributions, such as *benevolence* and *dominance* of the interpersonal circumplex; the three dimensional evaluations *valence, activity, potence* [17]; social and physiological attributions that were empirically obtained from free text voice descriptions [18]; previous work [19]; and aspects of longer-term interpersonal attraction [20]: *attractiveness, age, gender, interest, affect*, and *naturalness*.

The speech material employed consisted of 15 male German speakers with no marked accent or dialect, recorded in clean conditions as specified in [21]. Turns of a dialog of spontaneous speech (ordering a pizza), where individual speaker characteristics can be manifested, were selected for this study. The average speech duration across all speakers' samples was 19.6 s, with standard deviation 5.2 s.

In all, 33 normal-hearing participants (17 males, 16 females) completed the 28-item questionnaire for each voice. They ranged in age from 20 to 39 years (mean = 26.9, SD = 3.7) and their mother tongue was German. They used AKG K-601 headphones (diotic listening) to listen to the voice stimuli in an acoustically damped room. The participants confirmed that they did not recognize any of the speakers. Each individual session took 20 minutes approximately and the listeners were compensated for their participation. It has to be noted that the duration of completing our 28-item questionnaire for 15 voices is close to that of collecting ratings using the BFI-10 to measure the Big-Five traits in [22].

For comparison, the author of [11] employed speech material from 25 male German speakers, which was rated by 32 listeners on 15 adjective pairs. The study in [12] involved 64 Scottish speakers (32 male) and 10 adjective scales. Each listener (of a total of 320, performing the test in an uncontrolled listener environment) rated one scale only for each of the voices.

3. Test results and factor analysis

The intraclass correlation coefficient ICC(2,k) was calculated as an index of inter-rater reliability for each questionnaire item, using the 'psych' package in R. Inter-rater agreements ranged from .61 ("unaffektiert", in English: "unaffected") to .92 ("sicher", in English: "confident"). The average agreement was .83 (standard deviation = .07).

From the questionnaire results, two main findings can be emphasized. Firstly, the continuous ratings given to the questionnaire item with the antonyms "sympathisch"–"unsympathisch" (in English: "likable"–"dislikable" correlated strongly with the likability scores collected in [22] with Pearson $r = .74, p < .09$. In that previous study, scripted sentences uttered by the same male speakers as in this study were employed as speech stimuli (wideband-quality speech). The likability scores were directly given as an answer to the question (translated from German): "How likable do you find this voice over the telefon?"). Listeners indicated their ratings on a continuous scale with the antonyms "likable" and "dislikable" at its ends. The strong correlation found indicates high agreement between different listeners that rate different utterances from the same speakers and that, for the test of this paper, the other questionnaire items did not influence the ratings on the "likable"–"dislikable" item.

Table 1: *Retained items (German adjectives on the right of the antonym pairs and corresponding English translations) and standardized factor loadings.*

Scale (right adjectives)		Factor Loadings				
German	English	1	2	3	4	5
attraktiv	attractive	0.77				
angenehm	pleasant	0.75				
unsympathisch	dislikable	-0.70				
hässlich	ugly	-0.60				
sicher	confident	0.82				
unentschieden	indecisive		-0.71			
dominant	dominant		0.67			
gelangweilt	bored			0.76		
emotional	emotional			-0.74		
gleichgültig	indifferent			0.49		
ruhig	calm				1.00	
entspannt	relaxed				0.53	
inkompetent	incompetent					0.74
intelligent	intelligent					-0.69

And secondly, a two sample t-test has revealed statistically significant differences ($p < .05$) in ratings from different listeners' genders for the items "affektiert"–"unaffektiert", "jung"–"alt", and "dumm"–"intelligent" (the right adjectives in English are "unaffected", "old", and "intelligent", respectively). Hence, male listeners perceived the male speakers significantly more intelligent, more affected, and older than the female listeners did.

A factor analysis was performed using all listeners' answers. The number of factors was determined by Horn's parallel analysis. The subsequent exploratory factor analysis was conducted with `oblimin` rotation and minimum residual factoring method. Items were retained in case of loadings $> .5$ and no cross-loading $< .3$ (14 out of 28 remained). An additional factor analysis of the remaining items explains 61% of variance (see Table 1 for the final items and main loadings from the 2nd factor analysis). As factor scores, means of retained items are calculated. The five factors are named:

1. (social and physical) *attractiveness* ($\alpha = .85$)
2. *confidence* ($\alpha = .79$)
3. *apathy*($\alpha = .75$)
4. *serenity* ($\alpha = .79$)
5. *incompetence* ($\alpha = .74$)

4. Perceptual dimensions and attributed personality

The attributions found in Table 1 deliberately excluded the Big-Five personality taxonomy, since validated questionnaires already exist to assess the Big-Five traits. However, for the first impression (zero-acquaintance scenarios), validity and reliability are not sufficiently confirmed for all constructs, except for e.g. *Extraversion* [23]. Indeed, for reliable personality ratings, close friends provide the best participants [8, 9]. In addition, these Big-Five do not seem to be of utmost importance for social relationships of (in-person) first encounters [18], for which the two-factor model of competence and benevolence provide a meaningful start. Our factors attractiveness/benevolence, confidence, apathy/interest, serenity/calmness, and competence can be seen as a more detailed interpersonal description still in line with the two-factor model. The Big-Five traits, of course, pro-

Table 2: *Pair-wise Pearson correlations among attributed Big-Five personality traits [22] and the factors identified in this work.*

	attract.	confid.	apathy	seren.	incomp.
Extrav.	-0.02	0.37	-0.42	-0.21	0.05
Agreea.	0.53*	-0.55*	-0.29	-0.26	-0.35
Consci.	0.52*	0.18	-0.87***	-0.29	-0.74**
Neurot.	-0.32	-0.70**	-0.03	-0.62*	0.31
Open.	0.53*	-0.13	-0.94***	-0.53*	-0.60*

*$p < .05$; **$p < .01$; ***$p < .001$

vide a solid basis for interpersonal attributions contingent upon more relevant experience, not available in a passive listening test, such as evidence for task-related competence or for social relationships.

For the 15 male recordings in question, Big-Five ratings were collected in [22] using the BFI-10 in German [5]. The same spontaneous dialogs as in this work were presented to the listeners.

Due to the analogous listening-only situation, the person attributions of Section 3 are likely to be correlated with the Big-Five personality ratings at hand. Pair-wise correlations between the Big-Five perceived personality traits and the factor loadings prove this assumption (Table 2). Fortunately, all correlations are meaningful, in principle giving support to our questionnaire results. The highest correlations are found for the *apathy* dimension, indicating that persons perceived as apathetic, disinterested, or uninvolved are also thought of not being conscientious nor open-to-experience. Notably, *attractiveness* is positively correlated with agreeable, conscientious, and open-to-experience personalities. Although all these three personality dimensions refer to positive traits in social interaction, the correlation *attractiveness–extraversion* is not significant as in other studies [22, 24]. Of course, more data, especially for female speakers, are needed.

As usual, an oblique rotation was used for the factor analysis. The resulting cross-correlations are providing an additional hint that social attractiveness/likability is not identical to benevolence as often argued [25], but also comprised of competence (*attractiveness* correlates negatively with *incompetence*: $r = -.88$, $p < .001$). This can be in-line with [12], that detected a strong correlation between male vocal attractiveness and dominance. Further cross-correlations indicate a negative effect of *attractiveness–apathy* ($r = -.51$), and relations between *confidence–serenity* ($r = .54$), *apathy–serenity* ($r = .51$), and *apathy–incompetence* ($r = .59$). The latter result might be a unique situation-dependent effect, as displaying apathy in an interaction might also be interpreted as status difference instead of incompetence.

As mentioned before, likability scores to the same 15 males of this study were also collected in [22]. The Pearson correlation between these likability ratings (to wideband-quality speech with different content) and the factor scores of the first dimension of the present study results to be strong ($r = .76, p < .09$). Hence, it seems that the *attractiveness* dimension closely reflects continuous voice likability scores directly given on a single scale with the antonyms “likable” and “dislikable” at its ends.

5. Prediction of perceptual dimensions from speech features

In this section, we assess different speech features contributing to the automatic prediction of the five perceptual dimensions identified in this work. Besides, regression analyses are conducted with SVM (Support Vector Machines) with polynomial kernel, aiming at predicting each of the factors separately.

As speech features, the extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS), a set of 88 features [26], has been extracted using the openSMILE toolkit [27]. The eGeMAPS features have been proposed as a common baseline for evaluation of research in voice analysis (e.g. paralinguistics) and have been shown to offer performance comparable to that obtained with other large (6373 features) brute-force parameter sets for classification of affective categories [26].

The ‘caret’ package in R [28] was employed for our analyses. We computed the average feature importance for regression on each of the five perceptual dimensions of speaker attribution. For that, the 88 eGeMAPS features were assigned into different feature groups based on their perceptual similarity instead of the signal-based categorization given in [26]. For instance, average F_0 is relevant for pitch level perception, whereas jitter (originally in the same signal-based category as F_0) is correlated with the perception of noise and unsteady voice quality [29]. The perceptual categories used are:

- pitch level (1 feature)
- loudness level (2 features)
- speech rate (2 features)
- noise (3 features)
- timbre (spectral energy distributions, 14 features)
- special parameters dedicated to phonetic settings (9 features), e.g. Hammarberg index (differences between maximum amplitudes (in dB) in the LTAS spectral bins of 0–2 and 5–8 kHz)
- intonation contour (mean and sd of rising and falling F_0 slopes, 4 features)
- measures for the unvoiced segments (4 features)
- dynamics (spreads and deviations of other feature values, 49 features)

The importance of different speech features for regression was calculated by evaluating the relationship between each predictor and the outcome. This evaluation is done by fitting a loess smoother between the outcome and the predictor. The importance measure is an R^2 statistic calculated for this model against the intercept only null model [28]. The average importance measure was computed for each feature group and for each perceptual dimension and displayed in Figure 1.

It can be observed that intonation features are distinct for the *attractiveness* and *incompetence* dimensions. This outcome corroborates the finding in [30] that speakers with high variability in F_0 tend to be perceived as more competent and benevolent, and vice-versa. The averaged importance measures of different feature groups follow a similar trend for *attractiveness* and *incompetence*, which can be explained by the large cross-correlation of $r = -.88$ previously mentioned.

A large importance of pitch level (the mean logarithmic F_0 on a semitone scale averaged over the whole utterance) is manifested for *confidence*, *apathy*, and *serenity*. Related effects were also found in previous works [11, 12], where it was asserted that males with low mean F_0 are perceived as more dominant. Also the features (Alpha Ratio, Hammarberg Index, and spectral slopes from 0–500 Hz abd 500–1500 Hz) calculated over

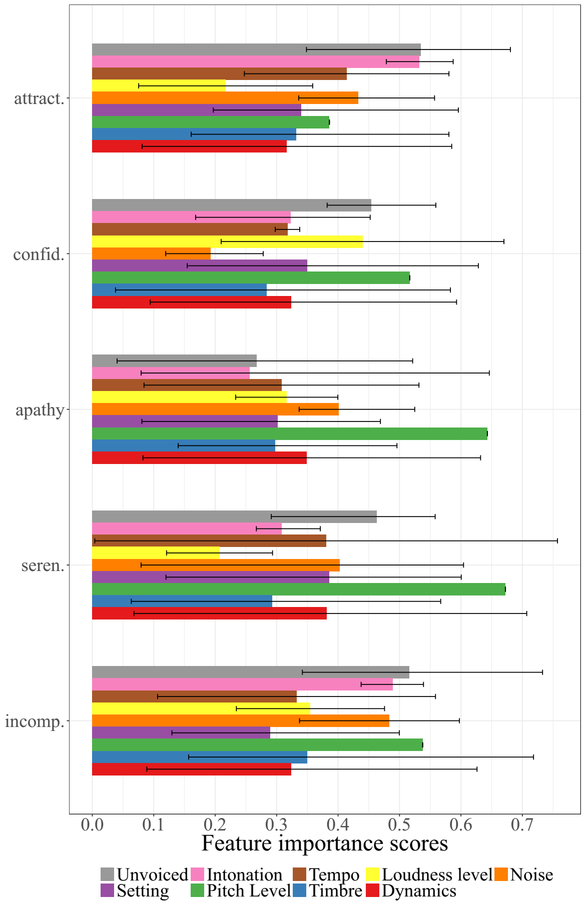

Figure 1: *Average importance of speech feature groups for regression of each perceptual dimension. The error bars represent percentile 5 and 95.*

unvoiced segments are outstanding in relation to the different speakers traits, except for *apathy*. Currently we do not have an explanation for this result.

The regression of each perceptual dimension was also conducted using the whole 88 feature set. Since only data from 15 voice samples are available, we opted for a leave-one-speaker-out cross-validation (LOSO-CV) scheme and performed SVM-based regression with polynomial kernel. As pre-processing, the extracted speech features were centered and scaled. The average performance is reported in Table 3 in terms of R^2 and root mean square error (RMSE), separately for each of the factors. The R^2 is calculated as the square of the Pearson's correlation between the observations and the model predictions. The RMSE results from squaring the residuals, averaging the squares, and taking the square root. The best-performing parameters (C, $scale$, and $degree$) leading to the highest R^2 have been selected [31].

It was noticed that, for *attractiveness*, *confidence*, and *apathy*, performances close to the best performance reached have also been obtained for higher degrees, with the regularization parameter $C = 10^{-3}$ or lower and with $scale = 10^{-3}$ or lower. The performance tended to decrease slightly with higher degrees. A polynomial kernel of $degree = 6$ offered the best performance for *serenity*. When experimenting with lower and higher degrees (from 1 to 9), performances with only a minor decrease with respect to $R^2 = 0.69$ were also found maintaining $scale = 10^{-4}$ and $C = 0.1$. For *incompetence*, performances around $R^2 = 0.60$ were obtained for $scale = 10^{-3}$ or lower and C between 10^{-4} and 10^{-1}. The performance tended to decrease for degrees higher than $degree = 2$.

Table 3: *SVM polykernel performance along with selected parameters.*

Factor	C	$scale$	$degree$	R^2	RMSE
attractiveness	10^{-4}	10^{-4}	1	0.68	0.42
confidence	10^{-4}	10^{-4}	1	0.65	0.58
apathy	10^{-4}	10^{-4}	1	0.58	0.53
serenity	0.1	10^{-4}	6	0.69	0.42
incompetence	0.1	10^{-3}	2	0.63	0.49

We are aware that more data would be needed to experiment with different machine learning algorithms with train/test data partitions and hence obtain more meaningful models for the perceptual dimensions. We leave such a study for future work.

6. Conclusions

In work we have identified, by means of a 28-item questionnaire of antonyms completed by 33 listeners for 15 male voices, five perceptual dimensions describing speaker attributions: 1. (social and physical) attractiveness; 2. confidence; 3. apathy; 4. serenity; and 5. incompetence. The relations between these traits and the Big-Five personality dimensions have been analyzed, and it has been discussed that the perceptual factors of this work can be valid for listening-only and zero-acquaintance scenarios.

Our findings reveal important feature sets for the prediction of our five perceptual dimensions. In concordance with previous literature, the relevance of intonation features is apparent in relation to speaker attractiveness and competence. Markedly, the mean F_0 has been found to be a primary feature for the detection of *confidence*, *apathy*, and *serenity*. Other features computed over unvoiced segments are relevant for the detection of the speakers traits of this work, except for *apathy*.

Using a baseline setup (LOSO-CV) based on the 88 eGeMAPS features and SVM regression with polynomial kernel, our regression analyses have shown performances varying from $R^2 = 0.58$ to $R^2 = 0.69$ for the different factors of person attribution. While this can be considered as an acceptable baseline performance, further analyses are required with additional substantial speakers' data in order to validate our results and to build useful models for the detection of perceptual dimensions. Specifically, we will conduct an analogous experiment using female speech, as we believe that the factors to describe female speakers' attributions may differ from those found for male speakers in this work. Our long-term goal is to derive an efficient listening test which should provide reliable descriptions of different speaker attributes, which can later be accurately detected from speech features.

7. Acknowledgments

This work has been supported by the German Research Foundation (DFG, Grant FE 1603/1-1 to Laura Fernández Gallardo).

8. References

[1] F. Burkhardt, H. R., and A. Batliner, "Application of Speaker Classification in Human Machine Dialog Systems," in *Speaker Classification I - Fundamentals, Features, and Methods*, C. Müller, Ed. Springer, 2007.

[2] J. Trouvain, S. Schmidt, M. Schröder, M. Schmitz, and W. J. Barry, "Modeling Personality Features by Changing Prosody in Synthetic Speech," in *International Conference: Speech Prosody*, 2006.

[3] K. R. Scherer, "Personality Inference from Voice Quality: the Loud Voice of Extroversion," *European Journal of Social Psychology*, vol. 8, pp. 467–487, 1978.

[4] A. Vinciarelli and G. Mohammadi, "A Survey of Personality Computing," *IEEE Transactions on Affective Computing*, vol. 5, no. 3, pp. 273–291, 2014.

[5] B. Rammstedt and O. P. John, "Measuring Personality in One Minute or Less: A 10-Item Short Version of the Big Five Inventory in English and German," *Journal of Research in Personality*, vol. 41, no. 1, pp. 203–212, 2007.

[6] G. An, S. I. Levitan, R. Levitan, A. Rosenberg, M. Levine, and J. Hirschberg, "Automatically Classifying Self-Rated Personality Scores from Speech," in *Annual Conference of the International Speech Communication Association (Interspeech)*, 2016, pp. 1412–1416.

[7] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore, "Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text," *Journal of Artificial Intelligence Research*, vol. 30, no. 1, pp. 457–500, 2007.

[8] D. W. Kolar, D. C. Funder, and C. R. Colvin, "Comparing the Accuracy of Personality Judgments by the Self and Knowledgeable Others," *Journal of Personality*, vol. 64, no. 2, pp. 311–337, 1996.

[9] B. S. Connelly and D. S. Ones, "An Other Perspective on Personality: Meta-Analytic Integration of Observers Accuracy and Predictive Validity," *Psychological Bulletin*, vol. 136, no. 6, pp. 1092–1122, 2010.

[10] J. S. Wiggins, P. Trapnell, and N. Phillips, "Psychometric and Geometric Characteristics of the Revised Interpersonal Adjective Scales (IAS-R)," *Multivariate Behavioral Research*, vol. 23, no. 4, pp. 517–530, 1988.

[11] M. Weirich, "Vocal Stereotypes," in *ISCA Tutorial and Research Workshop on Experimental Linguistics*, 2008, pp. 221–224.

[12] P. McAleer, A. Todorov, and P. Belin, "How Do You Say Hello? Personality Impressions from Brief Novel Voices," *PLoS One*, vol. 9, no. 3, p. e90779, 2014.

[13] A. Rosenberg and J. Hirschberg, "Acoustic/Prosodic and Lexical Correlates of Charismatic Speech," in *Annual Conference of the International Speech Communication Association (Interspeech)*, 2005, pp. 513–516.

[14] S. M. Smith and D. R. Shaffer, "Speed of Speech and Persuasion: Evidence for Multiple Effects," *Personality and Social Psychology Bulletin*, vol. 21, no. 10, pp. 1051–1060, 1995.

[15] E. Strangert and J. Gustafson, "What makes a good speaker? Subject Ratings, Acoustic Measurements and Perceptual Evaluations," in *Annual Conference of the International Speech Communication Association (Interspeech)*, 2008, pp. 1688–1691.

[16] B. Weiss and F. Burkhardt, "Voice Attributes Affecting Likability Perception," in *Annual Conference of the International Speech Communication Association (Interspeech)*, 2010, pp. 1934–1937.

[17] C. Osgood, G. Suci, and P. Tannenbaum, *The Measurement of Meaning*, ser. Illini Books, IB47. University of Illinois Press, 1957.

[18] B. Weiss, F. Burkhardt, and M. Geier, "Towards Perceptual Dimensions of Speakers' Voices: Eliciting Individual Descriptions," in *Workshop on Affective Social Speech Signals (WASSS)*, 2013.

[19] B. Weiss and S. Möller, "Wahrnehmungsdimensionen von Stimme und Sprechweise," in *Elektronische Sprachsignalverarbeitung (ESSV)*, 2011, pp. 261–268.

[20] E. Aronson, T. D. Wilson, and R. M. Akert, *Social Psychology*, 7th ed. Prentice Hall, 2009.

[21] L. Fernández Gallardo, "Recording a High-Quality German Speech Database for the Study of Speaker Personality and Likability," in *12. Tagung Phonetik und Phonologie im deutschprachigen Raum*, 2016, pp. 43–46.

[22] L. Fernández Gallardo and B. Weiss, "Speech Likability and Personality-based Social Relations: A Round-Robin Analysis over Communication Channels," in *Annual Conference of the International Speech Communication Association (Interspeech)*, 2016, pp. 903–907.

[23] B. Schuller and A. Batliner, *Computational Paralinguistics – Emotion, Affect and Personality in Speech and Language Processing*. Chichester: Wiley, 2014.

[24] 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," *Computer Speech & Language*, vol. 29, no. 1, pp. 100–131, 2015.

[25] M. Argyle, *Bodily Communication*. New York: Methuen, 1988.

[26] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. André, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan, and K. P. Truong, "The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing," *IEEE Transactions on Affective Computing*, vol. 7, no. 2, pp. 190–202, 2016.

[27] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent Developments in OpenSMILE, the Munich Open-Source Multimedia Feature Extractor," in *ACM Multimedia (MM)*, 2013, pp. 835–838.

[28] M. Kuhn. Contributions from J. Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, the R Core Team, M. Benesty, R. Lescarbeau, A. Ziem, L. Scrucca, Y. Tang, C. Candan and T. Hunt, *caret: Classification and Regression Training*, 2016, r package version 6.0-73. [Online]. Available: https://CRAN.R-project.org/package=caret

[29] J. Kreiman and B. R. Gerratt, "The Perceptual Structure of Pathologic Voice Quality," *Journal of the Acoustical Society of America*, vol. 100, no. 3, pp. 1787–1795, 1996.

[30] G. B. Ray, "Vocally Cued Personality Prototypes: An Implicit Personality Theory Approach," *Journal of Communication Monographs*, vol. 53, no. 3, pp. 266–276, 1986.

[31] A. Karatzoglou, D. Meyer, and K. Hornik, "Support Vector Machines in R," *Journal of Statistical Software*, vol. 15, no. 9, pp. 1–28, 2006.