Perceived Interpersonal Speaker Attributes and their Acoustic Features

Laura Fernández Gallardo, Benjamin Weiss

Quality and Usability Lab, Telekom Innovation Laboratories, TU Berlin, Germany

(laura.fernandezgallardo|benjamin.weiss)@tu-berlin.de

Abstract

This work investigates acoustic correlates of interpersonal speaker characteristics. Previous research has studied the effects of speech parameters on listeners impressions of e.g. attractiveness, charisma, and personality. Differently, we have derived five new factors of perceived speaker characteristics by employing the newly compiled Nautilus Speaker Characterization Corpus, of 300 German speakers. These factors are warmth, attractiveness, confidence, compliance, and maturity. Analyses of feature importance have revealed that pitch and other spectral features directly extracted from the signals correlate most with the target factor scores. With the gained knowledge of relevant features for each speaker trait, prediction models will be trained and tested in future work pursuing satisfactory speaker characterization performance.

Introduction

Prosodic features play a crucial role in transporting and evoking impressions of speakers' traits and states, such as personality or emotion. These paralinguistic aspects affect communication, behavior and relationship development. In order to identify prosodic features relevant for the attribution processes on the listeners' side, the relationship between acoustics and subjective perceptions has to be studied.

Previous investigations examining the acoustic cues that influence listeners' perception of speakers have either manipulated voice parameters, e.g. fundamental frequency (F0) (Apple et al. 1979), F0 variance (Brown et al. 1974), formant dispersion (Feinberg et al. 2005), intensity (Robinson and McArthur 1982), or speech rate (Brown et al. 1974), or directly employed own speaker recordings, as Collins (2000) and Zuckerman and Miyake (1993). Most of these studies have concentrated on impressions of vocal attractiveness (Zuckerman and Miyake 1993), likability, charisma, or personality (Apple et al. 1979).

Commonly, the two-dimensional personality aspects – *Competence/Dominance* and *Warmth/Benevolence* – of the well-known interpersonal circumplex have been examined (Brown et al. 1974 and Ray 1986). It has been shown by McAleer et al. (2014) that perceived attractive male speakers are also seen as competent and warm (likeable and trustworthy) while, for female speakers, attractiveness correlates with warmth but not with dominance judgements.

In our research, we address the relations between acoustic speech features and novel perceptually assessed human traits. We have employed speech data and labels from the Nautilus Speaker Characterization (NSC) Corpus[1] (Fernández Gallardo and Weiss); a newly recorded speech database of 300 German speakers (126 males and 174 females). By conducting factor analyses, we have identified dimensions of perceptive speaker attributions that encompass those of the interpersonal circumplex, attractiveness, and others. Based on factor scores of these dimensions, the importance of speech features for the automatic detection of speaker traits has been determined.

Delving into the relationship between speech signal parameters and perceived speaker characteristics provides valuable insights that contribute to automatic speaker characterization. There is a recent interest in predicting users' traits by analyzing their voices for e.g. call centers or personal companions. The human-machine speech dialog systems can then be adapted to the inferred user's preferences and behaviors to generate higher user acceptance (Berg 2014).

NSC Corpus

NSC Speech Recordings

The NSC corpus was recorded in 2016/2017 at the Quality and Usability Lab of the Technische Universität Berlin, Germany. Clean microphone recordings were made in the "Nautilus" acoustically isolated room, which gives name to this database. Conversational speech was elicited

[1] freely available for scientific research at the CLARIN repository: hdl.handle.net/11022/1009-0000-0007-C05F-6

from 300 German speakers (126 males and 174 females, aged 18 to 35 years, no marked accent). Additionally, spontaneous neutral and emotional speech utterances and questions were produced by the speakers interacting with their interlocutor (a recording assistant). Further details are given in (Fernández Gallardo and Weiss).

Four scripted and four semi-spontaneous dialogs were elicited from the speakers, simulating telephone call inquiries where the speakers played the client's role and the interlocutor the agent's role. The conversational scenarios followed were extracted and adapted from the Short Conversation Tests of ITU-T Recommendation P.805 (2007).

NSC Labels: Speaker Characteristics

One of the semi-spontaneous dialogs, in which the speakers' task was to order a pizza, was labeled for each speaker by external raters according to 34 interpersonal speaker attributions, such as likable, attractive, competent, childish, etc.

A semantic differential scale questionnaire (the Speaker Characteristics (SC)-Questionnaire (Fernández Gallardo and Weiss)) was employed to this end. The questionnaire items are based on previous research on interpersonal traits (Wiggins et al. 1988 and Jacobs and Scholl 2005); the three dimensional evaluations *valence*, *activity*, *potence* (Osgood et al. 1957); frequent social and physiological attributions (Weiss et al. 2017); and aspects of longer-term interpersonal attraction (Aronson et al. 2009). A first version of this questionnaire was validated by Weiss and Möller (2011), and later applied by Fernández Gallardo and Weiss (2017) employing only a small set of 15 male voices (from the NSC corpus).

In all, 114 naïve and normal-hearing individuals (70 males, 44 females) have participated in the rating procedure. They wore Shure SRH240 headphones (diotic listening, frequency range 20-20,000 Hz). The stimulus files had 44.1 kHz sampling frequency and a mean duration of 23.0 s, standard deviation 3.3 s. On average, each of the raters completed the SC questionnaire for 23.2 female and 16.4 male speakers. As a result, each speaker has been rated, on average, 15.1 (SD = 1.2) times on 34-dimensional continuous scales of interpersonal characteristics.

Factor Analysis

Using the NSC corpus labels of speaker characteristics, an exploratory factor analysis has been conducted for male and for female speakers separately. The number of factors was determined by Horn's parallel analysis. The *oblimin* rotation and minimum residual factoring method have been applied using the *psych* package in R.

Items were retained when main loading <= .5 & (main loading - cross-loading) <= .2. An additional factor analysis of the remaining items explains 58% and 56% of variance for male and for female speakers, respectively. Cronbach's alphas have been examined and some items have then been removed to reach the maximum internal consistency possible for each factor. The final Cronbach's alpha values range from .89 to .71.

Five factors (speaker attributes) have been found, which are named (for male speech):

1. warmth
2. attractiveness
3. confidence
4. compliance
5. maturity

For female speech, the same dimensions are found although in a slightly different ordering: *compliance* and *confidence* are, respectively, the 3rd and the 4th dimension for female speakers.

The final items and corresponding main loadings are presented in Table 1 and Table 2, for male and female speakers, respectively.

The factor scores for each dimension, used for the analysis of feature importance of this work, have been obtained by an average across raters of the z-scored ratings, weighted by the loadings for each retained item (Fernández Gallardo and Weiss).

Feature Importance

The extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS) (Eyben et al. 2016), of 88 numeric features, have been extracted using the openSMILE toolkit (Eyben et al. 2013). The eGeMAPS set has been compiled by a group of experts in paralinguistics. This minimalistic feature set has been shown to offer an emotion classification performance comparable to that obtained with larger (6373 features) brute-force parameter sets.

In the following, the analyses of feature importance have been conducted separately for male and for female speakers.

By employing the *caret* package in R, the features were preprocessed with a Box-Cox transformation, centered, and scaled. Afterwards, collinear features were excluded based on the variance inflation factor (VIF) (Naimi et al. 2014) calculated with the *vifstep* function of the *usdm* package in R and the recommended threshold value of 10.

Different filter methods to calculate feature importance have been explored using the R package *mlr*. This can be seen as a previous step to

the automatic detection of the factor scores that represent perceived speaker attributes. The computation of feature importance enables feature ranking and the selection of a feature subset for the regression task (beyond the scope of this paper). The feature importance values have been calculated with the methods: '*carscore*', '*cforest.importance*', '*linear.correlation*', '*mrmr*', '*randomForest.importance*','*randomForestSRC.rfsrc*', and '*rank.correlation*'.

The Kendall's coefficient of concordance has been computed as an index of interrater reliability of the feature importance ordering derived by the eight different methods tested. Looking for the highest mean Kendall's agreement of each methods with the rest for male and for female speech, for the five dimensions, has led to choosing linear correlation values as reliable indicators of feature importance. The importance values are also straightforward to interpret: the Pearson correlation between feature and factor scores.

The positive or negative correlation between each feature and each speaker attribute can now be investigated. Only the top-performing features are presented.

Male Speakers

- Contributing to higher *warmth*: Higher F0 range (ρ=.52), higher mean spectral slope 0-500Hz (ρ=.27), lower sd of F1 (ρ=-.23) and F2 (ρ=-.23) frequencies.
- Contributing to higher *attractiveness*: Higher F0 range (ρ=.35), higher sd of Hammarberg Index (ρ=.23), lower mean length of unvoiced segments (ρ=-.22), higher sd of F0 (ρ=.21), lower median F0 (ρ=-.20).
- Contributing to higher *confidence*: Lower median F0 (ρ=-.29), higher F0 range (ρ=.25).
- Contributing to higher *compliance*: Lower sd length voiced segments (ρ=-.24), lower sd of F1 frequency (ρ=-.23), lower sd of falling slope for loudness (ρ=-.20), higher F1 (ρ=.19) and F2 (ρ=.19) bandwidth.
- Contributing to higher *maturity*: Lower median F0 (ρ=-.42), higher sd of F3 frequency (ρ=.33), higher sd of F3 bandwidth (ρ=-.31).

Currently, we do not have explanations for many of the relationships found, since features are viewed in isolation. Further analyses of feature combinations can presumably offer more conclusive justifications of our results.

That notwithstanding, there are some clear observations that corroborate previous research on voice characteristics. Lower and highly variable pitch frequency seems to play a major role displaying attractiveness and confidence in male speech, as also shown by Zuckerman and Miyake (1993) and Jones et al. (2010). The melodious voice resulting from variable pitch also indicates higher warmth (a friendly attitude). Related to these findings, Brown et al. (1974) showed that decreased intonation and increased F0 caused the speakers to be rated as less benevolent and less competent. Also, Apple et al. (1979) indicated that males with high-pitch voices are perceived as less truthful, less persuasive, weaker, and more nervous. Lower median F0 also correlates with maturity, as also found by Collins (2000). The correlation between the sd of Hammarberg Index and attractiveness suggests that the dynamics in the energy distribution in the spectrum contributes to high perceived attractiveness.

Female Speakers

- Contributing to higher *warmth*: Higher F1 frequency (ρ=.43), higher F0 range (ρ=.37), higher sd of F2 bandwidth (ρ=.29), higher sd of spectral flux (ρ=.22), lower sd of F1 frequency (ρ=-.21), higher median F0 (ρ=.21).
- Contributing to higher *attractiveness*: Higher F1 frequency (ρ=.38), higher sd of F2 bandwidth (ρ=.30), lower sd of F1 frequency (ρ=-.23), higher F0 range (ρ=.22), lower mean spectral slope 0-500Hz (ρ=-.20).
- Contributing to higher *compliance*: Lower sd of F1 frequency (ρ=-.34), higher F1 frequency (ρ=.31), lower loudness range (ρ=-.30).
- Contributing to higher *confidence*: Higher sd of falling slope for loudness (ρ=.33), higher F0 range (ρ=.32).
- Contributing to higher *maturity*: Lower median F0 (ρ=-.47), higher mean mfcc4 (ρ=.46), lower F1 frequency (ρ=-.38), higher mean mfcc2 (ρ=.34).

For female speech, warmth, attractiveness, compliance, and "lower" maturity are cued by higher mean frequency of the first formant, which is also an indicator of a more careful and precise articulation. As also found for male speakers, higher pitch range signals higher warmth, attractiveness and confidence. With respect to these outcomes, Zuckerman and Miyake (1993) detected that female attractiveness correlated positively with articulation and negatively with monotonousness. Finally, lower pitch signals maturity, which is coherent with the decrease of females' F0 with age (Nishio and Niimi 2008).

Conclusions

Five speaker factors that are similar for both genders have been identified by employing the

NSC database (Fernández Gallardo and Weiss). These dimensions are: *warmth*, *attractiveness*, *confidence*, *compliance*, and *maturity*. They can be seen as perceptual dimensions that represent subjective attributions measured from observers' first impressions of speakers based on speech only.

Our following analysis suggests that pitch frequency and spectral features such as formants and mfccs are related to the perceived speaker traits to a larger extent than the other eGeMAPS features, in line with previous literature.

Future work will concentrate on examining different regression techniques combined with feature selection that yield the best automatic speaker characterization performance. The effects of degradations inserted by transmission channels (mainly bandwidth and codec) on the recognition of speaker attributes will also be examined.

Referenzen

Apple, W., Streeter, L. A., and Krauss, R. M. (1979). Effects of Pitch and Speech Rate on Personal Attributions. *Journal of Applied Social Psychology*, 37(5):715–727.

Aronson, E., Wilson, T. D., and Akert, R. M. (2009). *Social Psychology*. Prentice Hall, 7 edition.

Berg, M. M. (2014). *Modelling of Natural Dialogues in the Context of Speech-based Information and Control Systems*, volume 108 of *Dissertations in Database and Information Systems*. Akademische Verlagsgesellschaft in cooperation with IOS Press.

Brown, B. L., Strong, W. J., and Smith, B. L. (1974). Fifty-four Voices from Two: The Effects of Simultaneous Manipulations of Rate, Mean Fundamental Frequency, and Variance of Fundamental Frequency on Ratings of Personality from Speech. *The Journal of the Acoustical Society of America*, 55(2):313–318.

Collins, S. A. (2000). Men's Voices and Women's Choices. *Animal Behaviour*, 60:773–780.

Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., André, E., Busso, C., Devillers, L. Y., Epps, J., Laukka, P., Narayanan, S. S., and Truong, K. P. (2016). The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing. *IEEE Transactions on Affective Computing*, 7(2):190–202.

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

Feinberg, D. R., Debruine, L. M., Jones, B. C., and Perrett, D. I. (2005). Manipulations of Fundamental and Formant Frequencies Influence the Attractiveness of Human Male Voices. *Animal Behaviour*, 69:561–568.

Fernández Gallardo, L. and Weiss, B. The Nautilus Speaker Characterization Corpus: Speech Recordings and Labels of Speaker Characteristics and Voice Descriptions. *Submitted to International Conference on Language Resources and Evaluation (LREC 2018)*.

Fernández Gallardo, L. and Weiss, B. (2017). Towards Speaker Characterization: Identifying and Predicting Dimensions of Person Attribution. In *accepted for Interspeech*.

ITU-T Recommendation P.805 (2007). *Subjective Evaluation of Conversational Quality*. International Telecommunication Union, CH-Geneva.

Jacobs, I. and Scholl, W. (2005). Interpersonale Adjektivliste (IAL). *Diagnostica – Zeitschrift für Psychologische Diagnostik und Differentielle Psychologie*, 51(3):145–155.

Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C., and Vukovic, J. (2010). A Domain-Specific Opposite-Sex Bias in Human Preferences for Manipulated Voice Pitch. *Animal Behaviour*, 79(1):57–62.

McAleer, P., Todorov, A., and Belin, P. (2014). How Do You Say 'Hello'? Personality Impressions from Brief Novel Voices. *PLoS One*, 9(3):e90779.

Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K., and Toxopeus, A. G. (2014). Where is Positional Uncertainty a Problem for Species Distribution Modelling? *Ecography*, 37(2):191–203.

Nishio, M. and Niimi, S. (2008). Changes in Speaking Fundamental Frequency Characteristics with Aging. *Folia Phoniatrica et Logopaedica*, 60(3):120–127.

Osgood, C., Suci, G., and Tannenbaum, P. (1957). *The Measurement of Meaning*. Illini Books, IB47. University of Illinois Press.

Ray, G. B. (1986). Vocally Cued Personality Prototypes: An Implicit Personality Theory Approach. *Journal of Communication Monographs*, 53(3):266–276.

Robinson, J. and McArthur, L. Z. (1982). Impact of Salient Vocal Qualities on Causal Attribution for a Speaker's Behavior. *Journal of Personality and Social Psychology*, 43(2):236–247.

Weiss, B., Estival, D., and Stiefelhagen, U. (2017). Studying Vocal Perceptual Dimensions of Non-experts obtained from Speaker (Dis-)Similarities assessed by Direct Comparisons. *Acta Acustica united with Acustica*.

Weiss, B. and Möller, S. (2011). Wahrnehmungsdimensionen von Stimme und Sprechweise. In *Elektronische Sprachsignalverarbeitung (ESSV)*, pages 261–268.

Wiggins, J. S., Trapnell, P., and Phillips, N. (1988). Psychometric and Geometric Characteristics of the Revised Interpersonal Adjective Scales (IAS-R). *Multivariate Behavioral Research*, 23(4):517–530.

Zuckerman, M. and Miyake, K. (1993). The Attractive Voice: What Makes it So? *Journal of Nonverbal Behavior*, 17(2):119–135.