The Nautilus Speaker Characterization Corpus: Speech Recordings and Labels of Speaker Characteristics and Voice Descriptions

Laura Fernández Gallardo, Benjamin Weiss
Quality and Usability Lab, Technische Universität Berlin, Germany
Ernst-Reuter-Platz 7, 10587 Berlin, Germany
{laura.fernandezgallardo, benjamin.weiss}@tu-berlin.de

Abstract

The Nautilus Speaker Characterization corpus is presented. It comprises conversational microphone speech recordings from 300 German speakers (126 males and 174 females) made in 2016/2017 in the acoustically-isolated room *Nautilus* of the Quality and Usability Lab of the Technische Universität Berlin, Germany. Four scripted and four semi-spontaneous dialogs were elicited from the speakers, simulating telephone call inquiries. Additionally, other spontaneous neutral and emotional speech utterances and questions were produced. Interactions between speakers and their interlocutor (who also conducted the recording session) are provided in separate mono files, accompanied by timestamps and tags that define the speaker's turns. One of the recorded semi-spontanous dialogs has been labeled by external assessors on 34 interpersonal speaker characteristics for each speaker, employing continous sliders. Additionally, 20 selected speakers have been labeled on 34 naïve voice descriptions. The corpus labels permit to investigate the speech features that contribute to human perceptions and automatic recognition of speaker social characteristics and interpersonal traits.

Keywords: speech resource, corpus design, speech recordings, speech labeling

1. Motivation

This paper presents the Nautilus Speaker Characterization (NSC) Corpus[1], a new language resource that has been recently collected and labeled for the study of speakers' interpersonal characteristics. More specifically, we investigate the correspondence between acoustic parameters and speaker social characteristics and interpersonal traits, such as confidence, competence, and vocal attractiveness.

The NSC corpus has been designed to study how the subjective perception and automatic recognition of the speakers' social traits are affected by different degradations introduced by voice channel transmissions. Most of existing publicly available databases present speech segments that have already been transmitted/distorted through telephone channels and hence cannot be employed for our analyses. Their sampling frequency is not sufficient for the evaluation of super-wideband (SWB, 50–14,000 Hz) conditions, or the speech was recorded in noisy or uncontrolled environments. A detailed review is provided in (Fernández Gallardo, 2016). In contrast, the newly acquired NSC corpus presents:

- Clean speech recordings made in the acoustically-isolated *Nautilus* room (which gives name to this database) with the high-quality AKG C 414B-XLS microphone. The format of the recordings is audio/wav, 48 kHz, 16 bit, 1-channel.
- Speech from 300 (126 m, 174 f) native German speakers without marked regional dialect, aged 18 to 35 years.
- Human-human conversational speech and interactions. Recordings of four scripted dialogs, four semi-spontaneous dialogs, and spontaneous neutral and emotional speech statements and questions.
- Speakers' demographic information and self-assessed personality.
- For the 300 speakers, externally-assessed continuous numeric labels of 34 interpersonal speaker characteristics on one of the semi-spontaneous dialogs. For 20 selected speakers, also continuous numeric labels of 34 naïve voice descriptions for the same speech material.

Recording sessions of about 45 minutes have been conducted for each speaker individually by a recording assistant, who acted as interlocutor. When the 300 sessions were completed, we proceeded to collect labels given by external raters listening to semi-spontaneous dialog turns. This speech was evaluated in terms of interpersonal speaker characteristics (SC), such as likable, attractive, competent, childish, etc. An average of 15 external evaluators rated each speaker. They employed a 34-item semantic differential questionnaire, which we will refer to as SC-Questionnaire (Fernández Gallardo and Weiss, 2017a).

According to the two major dimensions determined by factor analysis on the SC ratings (*warmth* and *attractiveness*), a set of 20 "extreme" speakers has been selected to study their voice peculiarities. These have been evaluated by 26 listeners who completed the voice descriptions (VD)-Questionnaire, also a 34-item semantic differential rating scale, for each of the 20 voices.

Our long-term aim is to examine acoustic correlates of subjective speaker attributions and the influence of transmission channels for understanding human communication and behavior. Furthermore, being able to automatically predict

[1]The ISLRN of this corpus is 157-037-166-491-1. The data has been made available at the CLARIN repository: `hdl.handle.net/11022/1009-0000-0007-C05F-6` under the CLARIN ACA+BY+NC+NORED license (freely available for scientific research).

speakers' traits from speech features may assist the development of human-machine conversational systems, which could adapt to the detected user's attributes (Burkhardt et al., 2007; Berg, 2014).

The remainder of this paper is as follows. Section 2 describes the recorded speech and their tags, while Section 3 specifies the speech recording setup. Section 4 is devoted to present the collection of database labels and the factor analyses conducted for determining perceptual factors of speaker characteristics and of voice descriptions. This paper concludes with Section 5, which discusses possibilities of using the described NSC resource.

2. Speech Material

During the recordings, the speaker sat in the acoustically isolated room *Nautilus*, and the interlocutor in the office room *Belafonte* of the Quality and Usability Lab of the Technische Universität Berlin, Germany. Two females undertook the task of the interlocutor for 279 and for 21 sessions, respectively, and they also participated in other sessions as speakers.

Scripted and semi-spontaneous dialogs were held between the speaker and the interlocutor. In addition, spontaneous speech was recorded, which corresponds to neutral and emotional speech and questions that are not part of any proposed dialog and were casually uttered during the recording session when speaker and interlocutor interacted (open microphone setting).

For the scripted dialogs, the speakers were asked to read given dialog turns (scripts) as naturally as possible and maintaining the wording. Differently, for the semi-spontaneous dialogs, the speaker and the interlocutor followed a given conversational scenario, which had been extracted and adapted from the Short Conversation Tests of (ITU-T Recommendation P.805, 2007). Table 1 shows the topics of the recorded dialogs, which simulated telephone calls involving some inquiries. The speaker assumed the client's role, while the interlocutor played the role of a contact person or agent.

Tag	Description
1 (a,b,c,d)	Dialog 1: health insurance inquiry
2 (a,b,c)	Dialog 2: mobile phone rate plan inquiry
3 (a,b,c,d)	Dialog 3: car rental inquiry
4 (a,b,c,d)	Dialog 4: real state agency inquiry
5	Dialog 5: car rental booking
6	Dialog 6: pizza order
7	Dialog 7: book from the library
8	Dialog 8: doctor's appointment
d	(repeated) semi-spontaneous dialog turn
s	neutral spontaneous speech
e	spontaneous emotional excerpt
q	spontaneous question
f	spontaneous short feedback

Table 1: Topics of the recorded scripted dialogs (first block), semi-spontaneous dialogs (second block), recorded spontaneous events (third block), and their tags.

During the recording session, speakers spontaneously manifested some natural emotions (e.g. amusement, excitement, frustration of needing many turn repetitions), uttered questions, mentioned something related to the recording tasks, or provided short feedbacks (e.g. "aha", "ok", etc.) to indicate their understanding of the interlocutor's instructions. All events were recorded as well as the interlocutors' speech. After voice activity detection for the delimitation of speech segments, these were tagged with the dialog turn they correspond to, or as belonging to any of the mentioned events, as indicated in Table 1. We refer as *Interaction* to the three files provided together:

- one audio/wav file corresponding to the speaker's speech,
- one audio/wav file corresponding to the interlocutor's speech, and
- a csv file containing the tags and timestamps of the speaker segments.

These *Interactions* are provided for the semi-spontaneous and for the spontaneous speech.

Additional details on the NSC data, as well as the instructions given to the speakers and their consent form, dialog scripts and scenarios, etc. can be found in the NSC documentation.

3. Recoding Setup

As previously mentioned, the recruited speakers performed the recordings in the acoustically-isolated room *Nautilus*. The room's dimensions are 2.75 m x 2.53 m x 2.10 m, and RT60 = 0.08 s at 2 kHz. The approximate distance from the microphone to the speaker's mouth was 35 cm. The interlocutor sat in the adjacent room *Belafonte*, subject to background noises. She listened to the speaker, gave the pertinent instructions, and acted as dialog partner by using the headset Sennheiser HMD 46.

The hardware connections between the speaker's and the interlocutor's room are depicted in Figure 1. The speech signals were recorded using the software Cubase 4 with 48 kHz sampling frequency and 32-bit quantization.

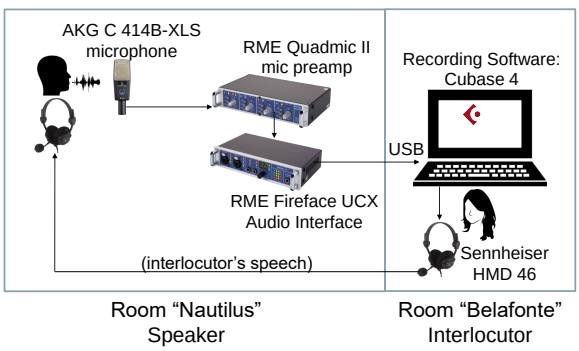

Figure 1: Diagram of device connections for the speech recording sessions.

4. Metadata and Speech Labels

Apart from the speech turn tags, also speakers' metadata and SC and VD labels are provided, as described in this section. The metadata was self-reported by each speaker at the end of the recording session.

Differently, the labels were collected by performing controlled listening tests with naïve normal-hearing external assessors (none of them participated as speaker). The semi-spontaneous speaker turns of Dialog 6 (pizza order) was chosen as speech stimulus (with mean duration 23.0 s, standard deviation of 3.3 s, range: 15.5 s–33.2 s). By controlling the speakers' dialog task, possible biases caused by disfluencies and different contents can be avoided for the estimation of speaker and voice attributes. At the same time, manifestations of speaker traits can still be perceived, as opposed to scripted speech.

4.1. Self-Reported Metadata

The socio-demographic data collected for each speaker comprise: age; gender; place of birth; chronological places of residence and duration of stay; place of birth of the mother; place of birth of the father; highest education level; educational background; main occupation; past occupations (if any); years of work experience (if any).

In addition, 273 (117 males and 156 females) out of the 300 speakers also completed a questionnaire for personality self-assessment (Rammstedt and John, 2007; Rammstedt and Danner, 2017) and self-assessed their vocal attractiveness using a continuous slider.

4.2. Interpersonal Speaker Characteristics

A group of 114 listeners (70 males and 44 females, 24.5 years old on average) participated in a series of listening tests to label perceived interpersonal speaker characteristics (SC), using the SC-Questionnaire with continuous sliders. 93 out of the 114 listeners spoke German as mother tongue, whereas the rest spoke other 10 different languages and claimed to have very good knowledge of German.

The SC-Questionnaire items (first two columns of Table 2) are based on previous research on interpersonal traits (Wiggins et al., 1988; Jacobs and Scholl, 2005); the three dimensional evaluations *valence, activity, potence* (Osgood et al., 1957); frequent social and physiological attributions (Weiss et al., 2018); and aspects of longer-term interpersonal attraction (Aronson et al., 2009). A first version of this questionnaire was validated (Weiss and Möller, 2011), and later applied (Fernández Gallardo and Weiss, 2017b) employing only a small set of 15 male voices. The version used for this work and its results are examined in (Fernández Gallardo and Weiss, 2017a).

The listeners first completed the SC-Questionnaire for a set of male voices and then for a set of female voices. Each assessor listened to and rated 16.4 male speakers and 23.2 female speakers on average, which results on an average of 15 x 34-dimensional continuous interpersonal ratings given by different listeners to each of the 300 speakers. They wore Shure SRH240 headphones (diotic listening, frequency range 20–20,000 Hz) and performed the test in a quiet office using a laptop and a mouse. They could listen to each speaker dialog as many times as they wished. Pauses were taken every 10 minutes approximately to avoid tiredness.

Using all continuous ratings, an exploratory factor analysis was conducted for male and for female speakers separately with `oblimin` rotation and minimum residual factoring method. The number of factors was determined by Horn's parallel analysis. Questionnaire items were retained when the main loading was greater than 0.5 and the difference between main loading and cross-loading exceeded 0.2. A second factor analysis was conducted on the remaining items, which explained 58% and 56% of data variance for male and for female voices, respectively. Cronbach's alphas were been examined and some items were removed to reach the maximum internal consistency possible for each factor.

We have identified five factors that are similar for both genders. These factors can be seen as perceptual dimensions that represent subjective attributions measured from observers' first impressions of speakers based on speech only. They are named (for male speech):

1. *warmth* (males: $\alpha = .88$, females: $\alpha = .89$)
2. *attractiveness* (males: $\alpha = .84$, females: $\alpha = .86$)
3. *confidence* (males: $\alpha = .78$, females: $\alpha = .80$)
4. *compliance* (males: $\alpha = .78$, females: $\alpha = .78$)
5. *maturity* (males: $\alpha = .76$, females: $\alpha = .71$)

The first, third, and fourth dimensions are described by the interpersonal circumplex (Wiggins et al., 1988; Jacobs and Scholl, 2005), while the second and fifth dimensions have their foundation on interpersonal attraction (Aronson et al., 2009) and age in speech (Weiss et al., 2018). 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 (Fernández Gallardo and Weiss, 2017a).

The questionnaire items associated with each factor and the corresponding loadings are presented in Tables 3 and 4 for male and for female speech, respectively. As factor scores, means of retained items are calculated, weighted by the item loadings.

4.3. Naïve Voice Descriptions

Considering the speakers' factor scores on the first two dimensions (warmth–attractiveness, with factor correlation of .59 and .63 for male and for female speech, respectively), a group of 20 "extreme" speakers have been selected: five male and five female speakers scoring lowest, and five female and five male speakers scoring highest in the warmth–attractiveness dimensional space. The subjective voice attributes of these speakers' speech has been thoroughly examined as indicated in this subsection, in order to better understand which acoustic cues are related to the different attributed speaker traits.

The 34-item semantic differential VD-Questionnaire (3rd and 4th columns of Table 2) is based on previous work (Scherer, 1974; Voiers, 1964; Fagel et al., 1983; Boves, 1984), was revised after (Weiss and Möller, 2011; Weiss et al., 2018), and then validated in (Weiss, 2016). This questionnaire has been completed by 26 assessors (13 males and 13 females, 26.6 years old on average, Ger-

SC: Antonyms (German)	SC: English translation
sympathisch / unsympathisch	likable / non-likable
unsicher / sicher	insecure / secure
unattraktiv / attraktiv	unattractive / attractive
verständnisvoll / verständnislos	sympathetic / unsympathetic
entschieden / unentschieden	decided / indecisive
aufdringlich / unaufdringlich	obtrusive / unobtrusive
nah / distanziert	close / distant
interessiert / gelangweilt	interested / bored
emotionslos / emotional	unemotional / emotional
genervt / nicht genervt	irritated / not irritated
passiv / aktiv	passive / active
unangenehm / angenehm	unpleasant / pleasant
charaktervoll / charakterlos	characterful / characterless
reserviert / gesellig	reserved / sociable
nervös / entspannt	nervous / relaxed
distanziert / mitfühlend	distant / affectionate
unterwürfig / dominant	conformable / dominant
affektiert / unaffektiert	affected / unaffected
gefühlskalt / herzlich	cold / hearty
jung / alt	young / old
sachlich / unsachlich	factual / not factual
aufgeregt / ruhig	excited / calm
kompetent / inkompetent	competent / incompetent
schön / hässlich	beautiful / ugly
unfreundlich / freundlich	unfriendly / friendly
weiblich / männlich	feminine / masculine
provokativ / gehorsam	offensive / submissive
engagiert / gleichgültig	committed / indifferent
langweilig / interessant	boring / interesting
folgsam / zynisch	compliant / cynical
unaufgesetzt / aufgesetzt	genuine / artificial
dumm / intelligent	stupid / intelligent
erwachsen / kindlich	adult / childish
frech / bescheiden	bold / modest

VD: Antonyms (German)	VD: English translation
klangvoll / klanglos	sonorous / flat
tief / hoch	low / high
nasal / nicht nasal	nasal / not nasal
stumpf / scharf	blunt / sharp
gleichmäßig / ungleichmäßig	even / uneven
akzentfrei / mit Akzent	accented / without accent
dunkel / hell	dark / bright
leise / laut	quiet / loud
knarrend / nicht knarrend	creaky / not creaky
variabel / monoton	variable / monotonous
angenehm / unangenehm	pleasant / unpleasant
deutlich / undeutlich	articulate / inarticulate
rau / glatt	coarse / not coarse
klar / heiser	clear / hoarse
unauffällig / auffällig	not remarkable / remarkable
schnell / langsam	quick / slow
kalt / warm	cold / warm
unnatürlich / natürlich	unnatural / natural
stabil / zittrig	stable / shaky
unpräzise / präzise	imprecise / precise
brüchig / fest	brittle / firm
unmelodisch / melodisch	not melodious / melodious
angespannt / entspannt	tense / relaxed
holprig / gleitend	bumpy / smooth
lang / kurz	long / short
locker / gepresst	lax / pressed
kraftvoll / kraftlos	powerful / powerless
flüssig / stockend	fluent / halting
weich / hart	soft / hard
professionell / unprofessionell	professional / unprofessional
betont / unbetont	emphasized/ not emphasized
sanft / schrill	gentle / shrill
getrennt / verbunden	disjointed / jointed
nicht behaucht / behaucht	not breathy / breathy

Table 2: 34 semantic-differential items of the SC (left) and VD (right) questionnaires. The 300 speakers have been labeled on the SC-Questionnaire items and the 20 selected "extreme" speakers on the VD-Questionnaire items. Continuous scales from 0 to 100 have been employed to evaluate every item.

Right adjective	Male SC factor loadings				
(translated)	warm.	attr.	conf.	comp.	matu.
hearty	.85				
affectionate	.84				
distant	-.76				
friendly	.59				
unsympathetic	-.58				
non-likable	-.52				
not irritated	.51				
attractive		.85			
ugly		-.79			
pleasant		.58			
interesting		.48			
secure			1.00		
indecisive			-.60		
submissive				.87	
cynical				-.71	
old					.82
childish					-.73

Table 3: Male SC factor loadings

Right adjective	Female SC factor loadings				
(translated)	warm.	attr.	comp.	conf.	matu.
hearty	.84				
affectionate	.84				
distant	-.78				
friendly	.56				
unsympathetic	-.49				
not irritated	.49				
non-likable	-.45				
attractive		.83			
ugly		-.81			
pleasant		.59			
submissive			.80		
cynical			-.72		
secure				.82	
indecisive				-.81	
childish					-.81
old					.68

Table 4: Female SC factor loadings

Right adjective (translated)	Male VD factor loadings *neg prof.	tens.	melo.	brig.
firm	-.78			
precise	-.72			
unprofessional	.65			
smooth	-.65			
shaky	.65			
halting	.64			
inarticulate	.61			
hard		.70		
pressed		.66		
relaxed		-.59		
warm		-.55		
jointed		-.50		
shrill		.49		
monotonous			-.78	
melodious			.65	
not emphasized			-.59	
bright				.87
high				.76
sharp				.58

Table 5: Male VD factor loadings

Right adjective (translated)	Female VD factor loadings *neg flue.	brig.	*neg prof.	shri.
shaky	.75			
firm	-.71			
smooth	-.67			
halting	.61			
high		.85		
bright		.79		
not coarse		.47		
hoarse		-.31		
unpleasant			.90	
inarticulate			.66	
unprofessional			.58	
shrill				.72
hard				.66
warm				-.52

Table 6: Female VD factor loadings

man as mother tongue), for each of the 20 selected "extreme" speakers using the same speech material (the pizza dialog).

Our procedure for factor analysis, conducted analogously as the one previously described, revealed four different dimensions for the description of the male and the female voices. The second factor analysis explained 54% and 53% of data variance for male and for female voices, respectively. For male speech, the dimensions found are:

1. *proficiency precision and fluency* (*negative, $\alpha = .87$)
2. *tension* ($\alpha = .79$)
3. *melody* ($\alpha = .83$)
4. *brightness* ($\alpha = .81$)

and, for female speech:

1. *fluency* (*negative, $\alpha = .81$)
2. *brightness* ($\alpha = .76$)
3. *proficiency precision* (*negative, $\alpha = .82$)
4. *shrillness* ($\alpha = .71$)

The gender difference in the dimensions found might be due to the small number of speakers tested. It can be speculated that *shrillness* is only manifested for female speech due to their generally higher pitch level compared to males. Tables 5 and 6 present the factor loadings calculated for each of the retained questionnaire items for male and for female speech, respectively. It has to be noted that items would load on positive factors tagged with **negative* with the opposite sign as the one indicated in the tables, e.g. the *precise* and *unprofessional* adjectives would load on positive *proficiency precision and fluency* for male speech with .72 and -.65, respectively.

We then examined the effects of speakers' warmth–attractiveness on the obtained VD factor scores. Conducted Wilcoxon rank-sum tests suggested that, for male speakers, *melody* and *brightness* factor scores differ significantly for perceived low warm–attractive speakers compared to perceived high warm–attractive speakers ($p < .01$ and $p < .05$, respectively). For female speakers, this statistical significant difference has been found for *fluency* ($p < .01$), *brightness* ($p < .05$), and *proficiency precision* ($p < .01$). These findings indicate the plausibility to classify perceived speaker traits based on speech features related to their voice descriptions. However, the voice descriptions of more speakers need to be analyzed in order to better determine the statistical effects between the SC and VD dimensions.

5. Conclusions and Outlook

In this paper we have presented the NSC corpus, which comprises clean conversational speech recordings from 300 German speakers and continuous numeric externally-assessed labels of speaker characteristics and voice descriptions.

The NSC corpus has been made freely available to the scientific community at the CLARIN repository, as mentioned before. The participating speakers gave their consent that "*the collected data will be exclusively used for scientific research and teaching activities. Accredited scientific institutions may access the data but not distribute them to third parties.*" (translated from German). It is foreseen that the data will also be available at the ELRA and LDC repositories.

The entire corpus material (50 GB of data) comprises 9192 minutes of speech (wav), the files employed as stimuli for SC- and VD-labeling (wav, 115 minutes), csv files with speakers' turn tags, listeners' ratings using the SC and VD questionnaires, SC and VD items–dimensions information, factor scores derived from the conducted factor analyses, speakers' metadata, and database documentation.

The collected data contributes to the investigation towards the detection of acoustic and linguistic cues that manifest subjective speaker social attributes. Following the Brunswik Lens Model revised by Scherer (1978), the features that can be extracted from the speech signal (e.g. pitch and formant frequencies, speech tempo, etc.) can be seen as "Distal Cues", whereas the collected VC-subjective labels

represent the "Proximal Percepts" that directly account for the final listeners' impressions of speakers (i.e. the SC dimensions identified in our analysis in Subsection 4.2.). The NSC data facilitates the research necessary to clarify the relationship between "Distal Cues" and "Proximal Percepts", which should lead to machines reaching the human performance in the attribution of speaker social characteristics.

The automatic detection of speaker interpersonal characteristics and traits is relevant to improve adaptive human-machine speech dialog systems. Speech and prosody production and conversational behavior in human-human interactions can be studied by analyzing speaker's and interlocutor's turns of semi-spontaneous and spontaneous speech. Based on recognized attributes from the users, systems should be able to adapt their dialog strategy and language generation mechanisms pursuing higher user acceptance (Berg, 2014).

The NSC data material may also be of interest to phoneticians and speech scientists requiring high-quality clean recordings in German.

6. Acknowledgments

This work has been supported by the German Research Foundation (DFG) [grant FE 1603/1-1].

7. Bibliographical References

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.

Boves, L. (1984). *The Phonetic Basis of Perceptual Ratings of Running Speech*. Dordrecht, Holland ; Cinnaminson, U.S.A. : Foris Publications.

Burkhardt, F., Huber, R., and Batliner, A. (2007). Application of Speaker Classification in Human Machine Dialog Systems. In C. Müller, editor, *Speaker Classification I - Fundamentals, Features, and Methods*. Springer, Berlin, Heidelberg.

Fagel, W. P. F., Van Herpt, L. W. A., and Boves, L. (1983). Analysis of the Perceptual Qualities of Dutch Speakers Voice and Pronunciation. *Speech Communication*, 1(4):315–326.

Fernández Gallardo, L. and Weiss, B. (2017a). Perceived Interpersonal Speaker Attributes and their Acoustic Features. In *Phonetik und Phonologie im deutschprachigen Raum (PundP13)*.

Fernández Gallardo, L. and Weiss, B. (2017b). Towards Speaker Characterization: Identifying and Predicting Dimensions of Person Attribution. In *Annual Conference of the International Speech Communication Association (Interspeech)*, pages 904–908.

Fernández Gallardo, L. (2016). Recording a High-Quality German Speech Database for the Study of Speaker Personality and Likability. In *Phonetik und Phonologie im deutschprachigen Raum (PundP12)*, pages 43–46.

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.

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

Rammstedt, B. and Danner, D. (2017). Die Facettenstruktur des Big Five Inventory (BFI). *Diagnostica*, 67:70–84.

Rammstedt, B. and John, O. P. (2007). 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*, 41(1):203–212.

Scherer, K. R. (1974). Voice Quality Analysis of American and German Speakers. *Journal of Psycholinguistie Research*, 3:281–298.

Scherer, K. R. (1978). Personality Inference from Voice Quality: the Loud Voice of Extroversion. *European Journal of Social Psychology*, 8:467–487.

Voiers, W. D. (1964). Perceptual Bases of Speaker Identity. *The Journal of the Acoustical Society of America*, 36:1065–1073.

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

Weiss, B., Estival, D., and Stiefelhagen, U. (2018). Studying Vocal Perceptual Dimension of Non-experts by Assigning Overall Speaker (Dis-)Similarities. *Acta Acustica united with Acustica*, 104:174–184.

Weiss, B. (2016). Voice Descriptions by Non-Experts: Validation of a Questionnaire. In *Phonetik und Phonologie im deutschprachigen Raum (PundP12)*, pages 228–231.

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.