Prosody, Phonology and Phonetics

Benjamin Weiss
Jürgen Trouvain
Melissa Barkat-Defradas
John J. Ohala *Editors*

Voice Attractiveness

Studies on Sexy, Likable, and Charismatic Speakers

Springer

Prosody, Phonology and Phonetics

The series will publish studies in the general area of Speech Prosody with a particular (but non-exclusive) focus on the importance of phonetics and phonology in this field. The topic of speech prosody is today a far larger area of research than is often realised. The number of papers on the topic presented at large international conferences such as Interspeech and ICPhS is considerable and regularly increasing. The proposed book series would be the natural place to publish extended versions of papers presented at the Speech Prosody Conferences, in particular the papers presented in Special Sessions at the conference. This could potentially involve the publication of 3 or 4 volumes every two years ensuring a stable future for the book series. If such publications are produced fairly rapidly, they will in turn provide a strong incentive for the organisation of other special sessions at future Speech Prosody conferences.

More information about this series at http://www.springer.com/series/11951

Benjamin Weiss · Jürgen Trouvain ·
Melissa Barkat-Defradas ·
John J. Ohala
Editors

Voice Attractiveness

Studies on Sexy, Likable, and Charismatic Speakers

Springer

Editors
Benjamin Weiss
Technische Universität Berlin
Berlin, Germany

Jürgen Trouvain
Saarland University
Saarbrücken, Germany

Melissa Barkat-Defradas
ISEM
Montpellier, France

John J. Ohala
International Computer Science Institute
Berkeley, CA, USA

ISSN 2197-8700 ISSN 2197-8719 (electronic)
Prosody, Phonology and Phonetics
ISBN 978-981-15-6626-4 ISBN 978-981-15-6627-1 (eBook)
https://doi.org/10.1007/978-981-15-6627-1

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

John Ohala passed away in August 2020, shortly before the publication of this volume. We dedicate this book to the memory of a great researcher.

Preface

At the Interspeech conference 2015, in Dresden, John (Ohala) asked Jürgen (Trouvain) what he thinks about organizing a special session on attractive voices, maybe for the next conference in this series. A former visiting researcher in Berkeley, Melissa (Barkat-Defradas), had already expressed some ideas on such an event on this topic. John has a long-standing interest in evolutionary aspects of speech and voice, Melissa works in an interdisciplinary research team on all kinds of aspects of evolution, and Jürgen has some background in paralinguistic characteristics of speech. At the same conference in Dresden, Jürgen introduced Benjamin (Weiss) to John with Benjamin as the optimal complement to this team since he has published several papers on social likability of voices.

It was then at Interspeech in Stockholm 2017, that we were able to organize the planned special session on voice attractiveness. We considered this event as the perfect setting for presenting research dealing with many aspects: perceived vocal preferences of men, women, and synthesized voices in well-defined social situations, acoustic correlates of voice attractiveness/pleasantness/charisma, interrelations between vocal features and individual physical and physiological characteristics, consequences for sexual selection, predictive value of voice for personality and for other psychological traits, experimental definition of esthetic standards for the vocal signal, cultural variation of voice attractiveness/pleasantness and standards, and also the link between vocal pathology and vocal characteristics. In Stockholm we agreed on a follow-up publication where the authors have more space than in a conference paper with its strict limitations. Moreover, also those colleagues could be reached that were not participants of this conference.

The special session was a success in our view. In total, we had nine accepted contributions. Authors from six papers of this session are also aboard in this volume. In addition to these, there are ten further contributions for this publication, having a total of seventeen papers when we add the introductory chapter. It is our belief that both collections, the nine conference papers, and the seventeen articles in this volume, can provide a useful overview on the state-of-the-art research on voice attractiveness, voice likability, and vocal charisma. We also hope that these studies

represent a fruitful fundament for further thoughts and investigations of an exciting field of speech and voice research.

As many book projects of this size, the editing process took longer than expected. This delay is mainly but note entirely due to health reasons of some of the editors. We would like to thank all authors for their patience and the publishing house for the provided support.

Berlin, Germany — Benjamin Weiss
Saarbrücken, Germany — Jürgen Trouvain
Montpellier, France — Melissa Barkat-Defradas
Berkeley, USA — John J. Ohala
April 2020

Contents

Editors and Contributors

About the Editors

Benjamin Weiss received his Ph.D. in 2008, in phonetics from Humboldt-University, Berlin. Since then he has extensively studied acoustic correlates of pleasant and likable voices, taking into account also speaking styles and conversational behavior in order to build quantitative models. He was visiting fellow at the University of Western Sydney and the University of Technology Sydney. In 2019, he completed his habilitation on human dialog and speech-based (multimodal) HCI. Since September 2020, he is an Associate Professor at the School of Intelligence, Hanyang University, Seoul.

Jürgen Trouvain received his Ph.D. in Phonetics in 2004, from Saarland University (Germany), where he works as a Senior Researcher and Lecturer at the Department of Language Science and Technology. His research fields include nonverbal vocalizations such as breathing and laughing, as well as non-native speech and phonetic learner corpora. He has acted as an organizer for several international conferences and workshops.

Melissa Barkat-Defradas obtained her Ph.D. in Forensic Linguistics at the University of Lyon, in 2000, and received the Young Researcher Award for her work in Automatic Language Identification. After a research fellowship at UC Berkeley, she joined the French National Centre for Scientific Research. She is now a full-time Researcher at The Institute of Evolutionary Sciences of Montpellier (France), where she actively contributes to developing interdisciplinary research by bridging the gap between experimental phonetics and evolutionary biology. She is particularly interested in the selective forces that may explain the emergence of articulated language in humans.

John J. Ohala is an Emeritus Professor of Linguistics at the University of California, Berkeley, and a Research Scientist at the International Computer Science Institute, Berkeley. He has had a major impact on the field of speech communication. His research interests focus on experimental phonology and phonetics and ethological aspects of communication, including speech perception, sound change, phonetic and phonological universals, psycholinguistic studies in phonology, and sound symbolism. He proposed an innovative ethological hypothesis, which unifies—via "the frequency code"—such diverse behavioral phenomena as the cross-language use of voice pitch for questions and statements, the systematic use of consonants, vowels, and tones in sound symbolical vocabulary, the "smile," and sexual dimorphism of the vocal anatomy in adult humans.

Contributors

Anita Auszmann Department of Head and Neck Surgery and Linguistics, University of California, Los Angeles, CA, USA

Molly Babel Department of Linguistics, University of British Columbia, BC, Canada

Melissa Barkat-Defradas Institut des Sciences de l'Evolution de Montpellier, University of Montpellier, Centre National de la Recherche Scientifique, Institut pour la Recherche et le Développement, Ecole Pratique des Hautes Etudes – Place Eugène Bataillon, Montpellier, France

Timo Baumann Universität Hamburg, Language Technology Group, Hamburg, Germany

Pascal Belin Institut de Neurosciences de La Timone, CNRS et Aix-Marseille Université Département de Psychologie, Université de Montréal, Montreal, Canada

Geke Boomsma Utrecht institute of Linguistics, Utrecht University, Utrecht, The Netherlands

Hans Rutger Bosker Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; Psychology of Language Department, Donders Institute for Brain Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

Alexander Brem Innovation and Technology Management, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

Felix Burkhardt audEERING GmbH, Berlin, Germany

Leigh Clark School of Information, & Communication Studies, University College Dublin, Dublin, Ireland; Computational Foundry, Swansea University, Swansea, UK

Benjamin R. Cowan School of Information, & Communication Studies, University College Dublin, Dublin, Ireland

Bruce R. Gerratt Department of Head and Neck Surgery, University of California, Los Angeles, CA, USA

Julia Hirschberg Columbia University, NYC, New York, NY, USA

Jody Kreiman Department of Head and Neck Surgery and Linguistics, University of California, Los Angeles, CA, USA

Grant McGuire Department of Linguistics, University of California Santa Cruz, Santa Cruz, CA, USA

Jan Michalsky University of Oldenburg, Oldenburg, Germany

Oliver Niebuhr Mads Clausen Institute, Centre for Electrical Engineering, University of Southern, Odense, Denmark

Yasunari Obuchi School of Media Science, Tokyo University of Technology, Hachioji, Tokyo, Japan

Abdulmalik Ofemile English Department, FCT College of Education, Zuba, Abuja, Nigeria

Hugo Quené Utrecht institute of Linguistics, Utrecht University, Utrecht, The Netherlands

Michel Raymond Institut des Sciences de l'Evolution de Montpellier, University of Montpellier, Centre National de la Recherche Scientifique, Institut pour la Recherche et le Développement, Ecole Pratique des Hautes Etudes – Place Eugène Bataillon, Montpellier, France

Andrew Rosenberg Google LLC, NYC, New York, NY, USA

Heike Schoormann University of Oldenburg, Oldenburg, Germany

Rosario Signorello Laboratoire de Phonétique et Phonologie, CNRS & Sorbonne Nouvelle, Paris, France

Alexandre Suire Institut des Sciences de l'Evolution de Montpellier, University of Montpellier, Centre National de la Recherche Scientifique, Institut pour la Recherche et le Développement, Ecole Pratique des Hautes Etudes – Place Eugène Bataillon, Montpellier, France

Ilaria Torre Department of Electronic and Electrical Engineering, Trinity College Dublin, Dublin, Ireland

Jürgen Trouvain Saarland University, Saarbrücken, Germany

Romée van Erning Utrecht Institute of Linguistics, Utrecht University, Utrecht, The Netherlands

Benjamin Weiss Technische Universität Berlin, Berlin, Germany

Laurence White School of Education, Communication and Language Sciences, Newcastle University, Newcastle, UK

Chloe Willis Department of Linguistics, University of California Santa Barbara, Santa Barbara, CA, USA

Part I
General Considerations

Chapter 1
Voice Attractiveness: Concepts, Methods, and Data

Jürgen Trouvain, Benjamin Weiss, and Melissa Barkat-Defradas

Abstract This book comprises contributions on vocal aspects of attractiveness, social likability, and charisma. Despite some apparent distinct characteristics of these three concepts, there are not only similarities, but even interdependencies to be considered. This chapter introduces and regards the concepts studied, methods applied, and material selected in the contributions. Based on this structured summary, we argue to increase interdisciplinary and even holistic efforts in order to better understand the concepts for voice and speech in humans and machines.

Keywords Attractiveness · Charisma · Likability · Sexual selection · Interdisciplinary · Holistic view · Structured summary · Speech production · Speech perception

1.1 Introduction

Probably, everybody has an idea of the meaning or meanings of *attractive* and *attractiveness* on the one side, and of voice and speaker on the other. It is also likely that everybody has their own ideas, which voices sound attractive—either in general or in specific contexts. But these ideas show by no means homogeneous structures and similar definitions.

A book on voice attractiveness attracts researchers, be it as authors and/or readers, who look at this topic from different angles as the subtitle of this book indicates. A *sexy* speaker is not the same as a *likable* speaker, and a *charismatic* speaker is different

J. Trouvain
Saarland University, Campus C7.2, 66123 Saarbrücken, Germany
e-mail: trouvain@coli.uni-saarland.de

B. Weiss (✉)
Technische Universität Berlin, Ernst-Reuter-Platz 7, 10405 Berlin, Germany
e-mail: benjamin.weiss@tu-berlin.de

M. Barkat-Defradas
University of Montpellier, Place Eugène Bataillon cc065, 34090 Montpellier cedex 05, France
e-mail: melissa.barkat-defradas@umontpellier.fr

B. Weiss et al. (eds.), *Voice Attractiveness*, Prosody, Phonology and Phonetics,
https://doi.org/10.1007/978-981-15-6627-1_1

again. These differences of how attractiveness is considered are also reflected in the chapters of this book. Likewise, the definition of speaker and voice is heterogeneously used, too. For this reason, we first attempt to shed some light onto the diversity of concepts we face in the upcoming chapters.

There is a broad range of different methods used in the studies of this volume. Many perform experimental research to investigate aspects of production, acoustics, and perception of attractive speech. There are some studies with a focus on modeling of data with respect to attractiveness, whereas other studies review how speech technology can be applied taking the (missing) attractiveness of voices into account. The data types that were used in the studies of this volume also show a large span. They range from manipulations of monosyllabic stimuli over single words and sentences in controlled settings up to many minutes of spontaneous conversational speech. The recap of the diversity of methods and data in this collection is followed by some concluding remarks on the emerging field of voice attractiveness, a research field that attracts researcher from many disciplines.

1.2 Concepts

1.2.1 Voice, Speaker, and Speech

The contributions of this collection consider the *voice* and *voice attractiveness* in different ways. Voice is not only seen in a narrow sense where it refers only to glottal activity. Voice in a wider sense additionally includes supra-glottal activities such as tongue raising, pharyngeal constriction, nasality or lip spreading (Laver, 1980), so that for instance formants as acoustic correlates of supra-laryngeal resonances are taken into account. For several studies, prosody plays an important role, reflected by fundamental frequency (F0), intensity, pauses and duration from a suprasegmental point of view. Further, timing parameters refer to entrainment in dialogs.

Naively, one would not assume that a voice that is considered as "normal", "stereotypical" or "average" would correlate to attractiveness. Nevertheless, three papers of this volume look more closely to the acoustic parameters of the "mean" voice and its perception of attractiveness—partially with somewhat surprising results.

Kreiman et al. (this volume) show that listeners differ regarding the question of what it means for a voice to sound "normal". There seem to be individual, rather consistent, strategies to label how normal or not normal a voice sounds. In their study, listeners assessed a wide range of one second samples of female speakers. From several acoustic parameters, the most relevant for explaining some amount of variance in the labels are fundamental frequency and its variation, as well as the first two formants, but not others that are typically associated with voice quality. However, the authors could not find a simple or generally valid answer, the situation is rather complex because several factors like the listener, the context, the purpose of the judgment, and of course the individual voice have to take into account.

The topic of recalling a voice from memory, an everyday task for everybody of us, is analyzed in Babel et al. (this volume). They show in a set of experiments with monosyllabic words as stimulus material that subjective stereotypicality and attractiveness affect the performance to remember a voice. Overall, they found support for the statement that less stereotypical voices and less attractive voices were better memorized.

Belin (this volume) reports of findings of experiments where identical short syllables of multiple voices of the same sex were averaged. The more voices were averaged the "speakers" of the averaged voice samples were perceived as more and more attractive. (similar to a visual effect concerning face attractiveness). Obviously, the main responsible factors for this effect are the reduced "distance-to-mean" for differences between F0 and the first formant, and an increased "texture smoothness" reflected by a raised harmonics-to-noise ratio.

There are also studies with stimuli to be rated that are longer than just one syllable or just one second. These studies concentrate more on speech prosody. Quené et al. (this volume), for instance, control for tempo and F0 in stimuli sentences, and Bosker (this volume) analyzed amplitude modulation in authentic speech samples. The review of charismatic speech of Rosenberg and Hirschberg (this volume) centers at prosody in all possible aspects, whereas, for instance, Weiss et al. (this volume) investigate acoustic parameters that reflect prosody (F0, intensity, rate), segmental properties (formants, spectral features) but also the voice in a narrow sense (shimmer, jitter, harmonics-to-noise ratio). These examples show that the vocal part in voice attractiveness can be referred to very different aspects of voice and speech when performing research in this field.

1.2.2 Sexual Selection and Voice Attractiveness

A sexy speaker can be seen as somebody who underlines her or his perceived sexual attractiveness—often unconsciously—with her or his voice and speech behavior. Though the voice is the privileged medium for interpersonal communication, it is not solely useful for conveying semantic information to other people. As a matter of fact, voice should also be regarded as a powerful social object, whose role is crucial in the context of human relationships. Indeed, by using oral communication, speakers are not only able to share their ideas and emotions, but they are also able to signal some reliable sociobiological features to their interlocutors such as sex, age, health, and social status, among others. There is a large body of scientific literature, for instance Scherer (1978), which describe the links between voice characteristics and personality traits, or the works by Laver and Trudgill (1979) and Bezooijen (1995), who studied voice as a social and cultural marker, or either still, Banse and Scherer (1996) whose work investigate how voice is used to express one's emotional state.

All of these authors, to name a few, have demonstrated that voice goes far beyond its primary linguistic function. Yet, interestingly, researches in Humanities mostly

tackled the topic of vocal function independently of any evolutionary considerations. However, as early as 1890, Darwin addressed the issue within the frame of sexual selection by drawing intriguing parallels between animal vocalizations and the human voice:

> The sexes of many animals incessantly call for each other during the breeding-season; and in not a few cases, the male endeavors thus to charm or excite the female. This, indeed, seems to have been the primeval use and means of development of the voice […]. When male animals utter sounds in order to please the females, they would naturally employ those which are sweet to the ears of the species; and it appears that the same sounds are often pleasing to widely different animals, owing to the similarity of their nervous systems, as we ourselves perceive in the singing of birds and even in the chirping of certain tree-frogs giving us pleasure. (Darwin, 1890, pp. 90–96).

Darwin's original idea according to which vocalizations allow the transmitter to attract females' attention and express his reproductive intentions make it legitimate to address the issue of human voice attractiveness in the specific context of human mating. As a matter of fact, as it is developed in the first contribution of Suire, Raymond, and Barkat–Defradas (this volume), it is reasonable to think that sexual selection—the mechanism which promotes biological and social traits that confer a reproductive benefit—has also intervened in the shaping of human vocal dimorphism; the attractiveness of a voice being a proxy, or a reinforcing signal, for other physical characteristics. By providing an overview of the research that lies at the crossroad of the human voice and evolutionary biology, the authors aim at demonstrating that sexual selection provides an interesting theoretical framework to understand the functional role of the human voice from an evolutionary perspective. Indeed, several studies have demonstrated the existence of a vocal attractiveness stereotype, which suggests that voice is an honest signal[1] of phenotypic quality in the same way as other physical features like, for example, the waist-to-hip ratio.[2]

Such an assumption raises the question of what makes a voice attractive? In their survey of the literature, Rosenberg and Hirschberg (this volume) examine the concept of vocal attractiveness itself. The authors consider the concept as highly context-dependent and discriminate between several types of attraction (i.e., political charisma, business leadership, nonsexual attraction and, last but not least, romantic desirability) each one of them being associated with specific articulatory, acoustic, and prosodic traits. They also show that though voice attractiveness is a complicated and exceptionally subjective phenomenon, evidence suggests some shared cross-cultural patterns that must have been shaped in the course of evolution by the selective pressure induced by the preferences of one sex for the vocal attributes of the other. The topic of vocal preferences has given rise to a large body of literature on the evolution of vocal preferences, which generally speaking, reveals that low-pitched

[1] Signals are traits that have evolved specifically because they change the behavior of receivers in ways that benefit the signaler. For example, peacock resplendent tail feathers are honest since they truly signal reproductive fitness of their bearer to the receiver.

[2] The waist-to-hip ratio (WHR) is the dimensionless ratio of the circumference of the waist to that of the hip. WHR correlates with health and fertility (with different optimal values in males and females).

masculine voices are universally preferred by women, such voices being perceived as related to a high quality phenotype. Conversely, men tend to prefer high-pitched feminine voices that are perceptually associated with youth and fertility at least in English. For more details of evolutionary mechanisms of attractive voices like mate choice see the systematic review of vocal preferences in humans by Barkat–Defradas, Raymond and Suire (this volume). Quené et al. (this volume) also confirm the expected pattern that men with lower-pitched voices tend to be rated as more attractive by (heterosexual) female listeners. They also reveal the importance of fast tempo in voice attractiveness evaluation. Indeed, their results based on manipulated speech show that the female raters judged masculine voices as less attractive if the F0 was artificially raised and the tempo decreased.

In their speed dating study, Michalsky and Schoormann (this volume) investigated the effects of perceived attractiveness and conversational quality on entrainment. In analyzing speed dating dialogs, prosodic disentrainment, in terms of pitch differences, is related to facial attractiveness for interlocutors of opposing sex. However, this result is inhibited by high conversational quality for females, and low conversational quality for males.

1.2.3 Likability and Social Attractiveness

A likable speaker is seen as somebody who underlines her or his perceived social attractiveness or pleasantness with her or his voice and speech behavior. There are several potential aspects that may constitute likability. For example, from the two of the most stable interpersonal concepts for unacquainted persons, benevolence (or warmth, communion) and competence (or agency, capability) (Abele, Cuddy, Judd, & Yzerbyt, 2008; Schaller, 2008; Fiske, Cuddy, & Glick, 2006), the first dimension (benevolence) is often assumed to resemble likability (DePaulo, Kenny, Hoover, Webb, & Oliver, 1987; Fiske et al., 2006; Argyle, 1988). However, liking-aversion may conceptually comprise the second dimension of competence as well (McCroskey & McCain, 1974), even in speech (Putnam & Street, 1984). Actually, there is much evidence from questionnaire analysis in a speech during dimension reduction that evaluative questionnaire items, such as "likable", can be apparent in both dimensions, benevolence and competence, or neither (Cuddy, Fiske, & Glick, 2008; Brown, Strong, & Rencher, 1973, 1985; Hart & Brown, 1974; Street & Brady, 1982; Weirich, 2010; Weiss & Möller, 2011). Given these empirical results, it can be argued that the so-called benevolence is just one possible but a very likely attribution to a person, which affects a speaker's social attractiveness, especially in a first impression.

Concerning voice acoustics, there are only few correlates of likability that show at least some robustness to changes in material, most notably increased pitch variability and tempo, while the results of average pitch reveal to be more complex, at least in German (Weiss et al., this volume).

While such results aim at correlates of averaged ratings on a scale, paired comparisons allow for a much finer measure of preference in likability. This method is, unfortunately, much more effort. Therefore, a crowd-based procedure is presented to collect such data efficiently, and it was used to train a model for predicting preferences of pairs of stimuli (Baumann, this volume).

In order to better take into regard the individual aspects of attractiveness, a method is presented that extracts overall voice attractiveness and listeners' preferences from paired comparisons, so that voices' likability can be estimated by the inner product of the two vectors of attractiveness and preferences (Obuchi, this volume).

1.2.4 Charisma and Leadership

A charismatic speaker is seen as somebody who underlines her or his perceived leadership, persuasive power, enthusiasm, and passion with her or his voice and speech behavior. Charisma is, just like likability, a social evaluation. However, likability typically refers to a dialogic situation, or in passive listening test, to the anticipation of a dialog—without any predefined difference in social status. In contrast to this, charisma is typically about an individual affecting a group of people, and thus implies some kind of social superiority. Charismatic people stand out, formally by social status or rank, or situationally by other's acknowledgment of their specialty. Therefore, the typical domains to study charisma in voice are speeches or talks of famous people, such as politicians and managers. A passionate and motivating speech by such people represents an often used, and sometimes even requested and anticipated, method of leadership. A discursive overview of what a charismatic voice actually is, can be found in Signorello (this volume).

The focus on public speeches and talks when dealing with charisma, complicates, on the one hand, differentiating between effects of a speech's presentation from those that originate in the fame, attributions, and social status. On the other hand, instead of relying on ratings in the laboratory, there a plenty of potentially valid indicators of charisma of those famous people including type of applause, (social) media reaction, and election results. For example, during a party conference of the German social democrats in 1995, the chairman was replaced by his vice-chairman—atypically early at this specific date—after an inspiring and enthusiastic speech of that vice-chairman. Given rather similar contents, sometimes even identical formulations, this outcome of the election was analyzed not regarding rhetorics, but speaking style instead (Paeschke & Sendlmeier, 1997). Such occurrences not only show that charisma is blended with power and leadership, but also exemplify the relevance of voice and speech for charisma. In this volume, the relevance of prosody and attire is studied for speeches of leading senior managers (Brem & Niebuhr, this volume). And in Bosker (this volume), a closer look on the modulation spectrum, which is related to speech rhythm, is taken for speeches from the US presidential campaign candidates Hillary Clinton and Donald Trump.

1.3 Methods

From a methodological perspective, we can divide studies on voice attractiveness in three fields. Investigations of the possible effects of different kinds of attractiveness and their vocal correlates are covered by *experimental research*. In addition to this research direction, *modeling* of processes how individual voices in audio samples attract listeners represents a further field of study. Finally, *technological applications* should be viewed as an own field of research in voice attractiveness.

1.3.1 Experimental Research

Human attractiveness is typically considered as a subjective concept. Therefore, experimental research is dominated by collecting explicit and implicit human ratings and decisions. The simplest methodological approach is to present stimuli and explicitly ask for ratings; on a scale if sequentially presented, or as a preference in the case of comparing stimuli. Such listening and ratings are, for example, conducted by Babel et al. (this volume). They collected a variety of subjective characteristics, among them perceptual similarity, applying a comparison of pairs of stimuli on a single scale, and perceptual attractiveness, collecting ratings in a sequential procedure for each stimulus individually. The latter method is also frequently used in the studies evaluated by Belin (this volume). Quené et al. (this volume) explicitly argue in favor of the sequential approach with absolute ratings instead of a forced preference choice of a direct comparison, as they want to avoid drawing attention to the signal manipulations they have conducted. There are various variants applied, often taken advantage of graphical computer interfaces, for example, to sort and assign short stimuli of a set to labels (Kreiman et al., this volume).

Instead of explicitly asking for measures of attractiveness, implicit measures can be attempted to collect, in order to avoid a social bias of the subjects. Such approaches comprise observations of social decisions, for example, counting the number of direct interactions in gaming or game-like tasks (Krause, Back, Egloff, & Schmukle, 2014). Other observations refer to the number of friends, or offspring (or explicitly asking to disclose the number of sexual partners). Such long-term or retrospective observations and surveys are, however, difficult to relate to specific traits, such as vocal characteristics.

1.3.2 Modeling

Quantitative modeling of subjective human ratings, such a sexual or social attractiveness, serves in principle two purposes. One is to describe the relations, e.g., correlations, found with parameters of interest in a given data set. Such a model could be a starting point for a prediction model, but does not provide explanatory power as

in a scientific theory. For the case of voice attractiveness, typical model parameters are acoustic or articulatory measures. Another purpose is to actually explain interdependencies between parameters and ratings in a quantitative way. However, in the latter case, the parameters chosen and the kind of relationship have to be confirmed by methodological means ensuring a causal relationship. Synthesizing or resynthesizing speech represents the most popular approach to control for the variables in question. It also aims at providing proof for a causal relationship. As the knowledge base is enhanced by empirical studies incrementally, each study might fulfill both purposes to some degree. For example, the linear models of social attractiveness of Weiss et al. (this volume) build on hypotheses drawn from several scientific methods in order to add evidence for acoustic-perceptual relations, but its main result is a simple data description.

Baumann (this volume), present a methodological approach, that does comprises not only the acoustic modeling part, but also a method to efficiently collect preference ratings for stimulus pairs. Such pairwise preferences for German spoken Wikipedia articles were acoustically correlated directly, and modeled as relative preferences by means of a recurrent neural network.

In a related approach, Obushi (this volume) collected pairwise preferences for a Japanese greeting phrase. The ratings are multidimensionally analyzed, taking into account the listeners' differences as well, and modeled by multiple acoustics features applying machine learning.

1.3.3 Technological Applications

Voice attractiveness can play an essential role in human-machine interaction (HMI) as two contributions in this volume show. There is a tendency that "people tend to attribute personality traits to computers and robots as if they were human agents" (Nass, Moon, Fogg, Reeves, & Dryer, 1995). That means that the human-sounding voices of talking and conversational computers can also be considered as personalized machines. In addition, machines can act for humans, for instance, when a speech synthesizer is used as a speech prosthesis for people who cannot clearly and fluently articulate anymore. From a view of listening to talking machines, we all know that it is most of the time rather boring and less interesting when faced with an artificial voice and synthesized speech, be it when street names are announced in car navigation or when interacting with a dialog system. For conversational agents, e.g., intelligent personal assistants, it is a particular challenge to show skills that are required for smooth dialogs that span aspects of timing up to common grounding. Thus, voice selection and voice modeling should be an integral part of the design in HMI tools. The paper collected in this volume are not empirical studies with existent systems but are reviews in which important thoughts are developed before experiments that test the usability of certain aspects of voice attractiveness are performed.

Torre and White (this volume) focus on the characteristics of a robot's voice in human-robot interaction. They are particularly interested in how vocal elements

can contribute to the impression of trustworthiness. They review studies in which a robot's voice was analyzed or manipulated, always with a particular view on trustworthiness. Naturalness and "machine-likeness", cognitive load, incongruity with the robot's behavior in general and the robot's appearance such as its size, gender, accent, and interaction context. Furthermore, they argue that the design of robot voices should come with an unambiguous appearance and function, because unrealistic expectations of robot performance in human users should be avoided.

The human evaluation in regard to different kinds of attractiveness represent immanent social and cognitive processes. Such evaluations are, however, not limited to other living persons. Instead, interactive systems, especially those using speech, are known to evoke similar processes (Reeves & Nass, 1996; Nass & Brave, 2005). And with the emergence of speech interaction with computers in the form of personal smartphone assistants, smart home devices, virtual persons, and human-like (social) robots, the users' appraisal of the verbal and nonverbal behavior of such interactive computers are receiving much attention.

One observation specific to anthropomorphic computers is the so-called "uncanny valley" effect. It describes an overall increase in familiarity (or attractiveness or likability) with increasing human-likeliness (or level of details) of the systems features and movements that is disrupted by a sudden decrease in familiarity close to perfect human-likeliness (Mori, 2012). This awkward or eerie feeling for a close to human, but obviously not natural synthesis is typically explained by a shift in reference from artificial to human and can be circumvented by reducing the level of human-likeliness or choosing an artificial metaphor (e.g., a puppet or cartoon) instead of a human. This effect is mostly studied for visual perceptions of the body and face of a robot or virtual person and their animated movements. However, in Clark (this volume), results for the evaluation of three linguistic strategies, politeness, relational work, and vague language are discussed in their usage for speech interfaces and their potential mismatch with the expectations in human users, and thus their potential to cause an uncanny valley effect.

One important sub-concept of social attractiveness is trust (McAleer, Todorov, & Berlin, 2014; Weiss, Wechsung, Kühnel, & Möller, 2015). In Torre and White (this volume) the effects of robot voices' gender, naturalness, prosody, and accent on trust perception in users are presented and systematized. Overall, there are effects, but they depend on the context and user group. For example, a regional accent showed an increased credibility to a standard accent when being knowledgeable, but the opposite in the case of being unknowledgeable.

1.4 Data

The material used in studies on voice attractiveness varies widely, from monosyllabic stimuli recorded in the lab to large extracts of authentic speech material that was not produced for research. This stylistic diversity is also reflected in the contributions

for this volume. Thus, it seems fair to separate three kinds of sources, controlled experimental data, naturalistic lab data, and natural field data "from the wild".

1.4.1 Controlled Experimental Data

One major source of the material stems from lab experiments, where new recordings are conducted for a specific purpose with already defined acoustic and perceptual analytic methods to be applied on. Such recordings are usually very short, for example (sustained) vowels, syllables or words. They can also not be considered as socially authentic, i.e., they do not aim to resemble real-life social communication situations. Due to its short duration, such material lacks major prosodic aspects, e.g., intonation contour or emphasis variation, as well as any natural situational grounding, affecting, e.g., speaking rate. Controlling for such aspects, however, allows to focus on topics like voice quality and person identification/similarity, while explicitly controlling for the just mentioned effects.

Examples of experimental data are Belin (this volume), who uses averaged short syllables of multiple voices, for which attractiveness ratings are collected. Kreiman et al., (this volume) analyzes steady state vowels (one second duration) regarding "normal" voice quality, whereas Babel et al., and Obuchi (both this volume) used single (monosyllabic, respectively multisyllabic) words for perception tests.

On some occasions, full sentences, or even a paragraph, are read by speakers in a lab with similar aims. The practical implications include potential laborious manual work to extract specific segments for analysis, and to take into account richer linguistic context, while the read speech style in a controlled environment allows to analyze not only segmental and micro-prosodic, but also macro-prosodic parameters. Therefore, it is not a coincidence to find a mixture of material types from experimental data in the cited literature for our topics that refer to social attributions and traits from speech (Suire et al.; Rosenberg & Hirschberg, both this volume). While some decisions on the material duration are made because of the costs inflicted by the prospective methods (see Sect. 1.3), other reasons to select material originate in the aspects under research.

The syllables used by Belin (this volume) were recorded in the lab, and subsequently post-processed to study the effect of acoustic averaging over speakers. Such a manipulation of speech recordings is another kind of experimental data. Manipulations comprise post-processing of the acoustic speech signal, as well as outright synthesis. Manipulated audio files can be in principle of any duration, but are considered here still as experimental data due to its similarity in careful and specific creation in a laboratory, but also due to the aim of controlling influencing factors—this time by means of inducing a controlled number of manipulations. There are different reasons for such manipulations, most importantly to verify analysis results with even more controlled material, producing stimuli for experiments which are hard or impossible to record, or to obtain speech signal qualities for the domain of computer speech.

The papers in the part on technological applications are good examples, as they all refer to studies in which manipulated or synthesized material, typically shorter utterances in a dialog, are used, or they argue to conduct those (Torre & White; Clark et al., both this volume).

1.4.2 Naturalistic Data Recorded in the Lab

While strictly controlled speech material from the laboratory is a foundation of basic research, there is always the aim to use naturalistic data in order to estimate the strength of effects for real-life situations and to study situational and dialogic aspects that cannot be simulated with—what we call—experimental data. Typically, this means to elicit naturalistic situations and thus also spontaneous material in the lab, often with the help of some supporting material. In contrast to the aforementioned controlled experiments, the lab recordings of naturalistic data are not controlled to the same degree. Here, experimenters aim to control a good acoustic quality, to initiate conversations, and possibly to instruct conversational tasks. That means that the linguistic and phonetic content is not (strictly) controlled for. However, very specific instructions and support material is often provided to support the subjects to elicit the situation, e.g., a game or task, but databases have been created with far less information provided (Schweitzer, Lewandowski, Duran, & Dogil, 2015).

For obtaining attractiveness ratings, Quené et al., (this volume) used sentences from spontaneous interview speech as stimuli that were manipulated. They also used visual data. The situation of speed dating was applied by Michalsky and Schoormann (this volume) to allow for studying the effects of prosodic entrainment in dialog. Simulated telephone conversations on pizza ordering from the Nautilus database, but post-edit to exclude the callee were used by Weiss et al. (this volume).

1.4.3 Data from the Wild

The last category of the material refers to recordings from real situations. Obtaining such data seems to be the easiest one on the first glance. However, it is often practically impossible to ensure sufficient quality and sufficient amount of material given the available resources, especially if there are requirements on the linguistic conditions to be included. In addition, there is often more information on the speakers required, which might be difficult to collect while or after recording, for example, additional physiological measures. Finally, there might be ethical reasons to avoid taking data from the wild.

In this collection, this kind of data was selected to solely study charismatic speakers. Bosker (this volume) selected speech fragments of c. 25 s from mass media recordings of US presidential debates. Brem and Niebuhr (this volume) used audio-visual data (video clips of charismatic management leaders). For natural data, this

kind of material is the least uncontrolled, as the speakers are not only professional, but also very aware of the fact of being recorded. Therefore, such field data might not always be considered as truly "wild", but of course, it is as natural as it can be when studying speeches of charismatic leaders.

Sometimes, it is not easy to assign data to one of the categories. For example, read Wikipedia articles used by Baumann (this volume) is comparable on the surface with other naturalistic speech paragraphs read in the lab, except for the varying recording quality. But still, the origin of this material is natural, as the speakers truly recorded themselves with the intention to be listened to by people interested in the Wikipedia articles.

1.5 Conclusions

The word "attractiveness" stems from Latin "ad trahere" and means "dragging or pulling to something". For our topic, people are dragged or pulled to the voice and vocal behavior of somebody else. This relationship unfolds in various dimensions: from sexuality and biology over social likability up to charisma and leadership. It is this diversity of voice attractiveness that we intended to cover in this book. It is our hope to raise awareness with this book for this diversity and the broad range of the various scientific fields involved.

What we see in the contributions to this volume is on the one hand a clear and intended separation of the above-mentioned concepts on the sexual, the likable, and the charismatic speaker. On the other hand, we recognize the interdependencies between the three concepts. The classical example is that a person perceived as beautiful is also regarded as a socially more attractive (Zuckermann & Driver, 1989).

In our view, we deal here with a contrast between simultaneous distinctive concepts that have not only mutual influences and mutual conditionality. We see a need for a unifying theory with respect to the concepts, but also the different methods and data used in the various scientific disciplines. Several contributions in this book provide useful suggestions for such a theory, which can be viewed as a starting point for a more systematic foundation to overcome the current limitations of knowledge.

As an example can serve the frequency code by Ohala (1984): Similarities between languages, cultures, and even species in the use and effect of F0 was argued to originate in biologically grounded separation between "smaller" and "larger" (vocal) individuals. This does not only reflect the sexual dimorphism in terms of sexual selection, but also social aspects of signaling and estimating relational power, submissiveness, even helplessness, and thus supports social roles and interaction. The universal systematic in F0 observed by Ohala concerns charisma, attractiveness, and likability alike. Following this road to connect biological and articulatory bases for acoustic and perceptual effects can be seen as one of the most important elements of a unifying theory.

Interestingly, we observe that *trust* occurs in many contributions and it seems to have an overarching character. Trust, obviously, represents a link between the

concepts of the sexual, the social, and the charismatic attractiveness, as it represents a positive attitude towards another. Trust may be considered as an immediate result of attractiveness, whatever the kind of attractiveness and social relation might be. Therefore, it is an important characteristic of human relationships, but also an important feature for Human-Computer Interaction.

References

Abele, A. E., Cuddy, A. J. C., Judd, C. M., & Yzerbyt, V. Y. (2008). Fundamental dimensions of social judgment. Editorial to the Special Issue. *European Journal of Social Psychology, 38*(7), 1063–1065.

Argyle, M. (1988). *Bodily Communication*. New York: Methuen.

Banse, R., & Scherer, K. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636.

Bezooijen, R. V. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, *38*, 253–265.

Brown, B. L., Strong, W. J., & Rencher, A. C. (1973). Perceptions of personality from speech: effects of manipulations of acoustical parameter. *Journal of the Acoustical Society of America*, *54*(1), 29–35.

Brown, B. L., Giles, H., & Thakerar, J. N. (1985). Speaker evaluation as a function of speech rate, accent, and context. *Language and Communication*, *5*(3), 207–220.

Cuddy, A. J., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in Experimental Social Psychology*, *40*, 62–149.

Darwin, C. (1890). *The Expression of the Emotions in Man and Animals*. London: John Murray.

DePaulo, B. M., Kenny, D. A., Hoover, C. W., Webb, W., & Oliver, P. V. (1987). Accuracy of person perception: Do people know what kinds of impressions they convey? *Journal of Personality and Social Psychology*, *52*(2), 303–315.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2006). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83.

Hart, R. J., & Brown, B. L. (1974). Personality information contained in the verbal qualities and in content aspects of speech. *Speech Monographs*, *41*, 271–380.

Krause, S., Back, M. D., Egloff, B., & Schmukle, S. C. (2014). Implicit interpersonal attraction in small groups automatically activated evaluations predict actual behavior toward social partners. *Social Psychological and Personality Science*, *20*, 671–679.

Laver, J., & Trudgill, P. (1979). Phonetic and linguistic markers in speech. In K. R. Scherer & H. Giles (Eds.), *Social Markers in Speech* (pp. 1–32). Cambridge: Cambridge University Press.

Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.

McAleer, P., Todorov, A. & Berlin, P. (2014). How do you say 'Hello'? Personality impressions from brief novel voices. *PLOS ONE 9*(3).

McCroskey, J., & McCain, T. (1974). *The Measurement of Interpersonal Attraction. Speech Monographs*, *41*, 261–266.

Mori, M. (2012). The uncanny valley. *IEEE Robotics and Automation 19*(2). Originally 1970, Translated by MacDorman, K.F. & Kageki, N., pp. 98–100.

Nass, C., & Brave, S. (2005). *Wired for Speech. How Voice Activates and Advances the Human-Computer Relationship*. MIT Press.

Nass, C., Moon, Y., Fogg, B., Reeves, B., & Dryer, D. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, *43*, 223–239.

Ohala, J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, *41*, 1–16.

Paeschke, A., & Sendlmeier, W. F. (1997). Die Reden von Rudolf Scharping und Oskar Lafontaine auf dem Parteitag der SPD im November 1995 in Mannheim –Ein sprechwissenschaftlicher und phonetischer Vergleich von Vortragsstilen. *Zeitschrift für Angewandte Linguistik*, *27*, 5–39.

Putnam, W. B., & Street, R. L. J. (1984). The conception and perception of noncontent speech performance: Implications for speech-accommodation theory. *International Journal of the Sociology of Language*, *46*, 97–114.

Reeves, B., & Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge: Cambridge University Press.

Schaller, M. (2008). Evolutionary basis of first impressions. In N. Ambady & J. J. Skowronski (Eds.), *First Impressions* (pp. 15–34). New York: Guilford Press.

Scherer, K. R. (1978). Personality inference from voice quality: The loud voice of extroversion. *European Journal of Social Psychology*, *8*(4), 467–487.

Schweitzer, A., Lewandowski, N., Duran, D., & Dogil, G. (2015). Attention, please!—Expanding the GECO database. In *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, paper 620.

Street, Jr. R. L., & Brady, R. M. (1982). Speech rate acceptance ranges as a function of evaluative domain, listener speech rate and communication context. *Communication Monographs 49*(4), 290–308.

Weirich, M. (2010). *Die attraktive Stimme: Vocal Stereotypes. Eine phonetische Analyse anhand akustischer und auditiver Parameter*. Saarbrücken: Verlag Dr. Müller.

Weiss, B., Wechsung, I., Kühnel, C., & Möller, S. (2015). Evaluating embodied conversational agents in multimodal interfaces. *Computational Cognitive Science*, *1*(6), 1–21.

Weiss, B., & Möller, S. (2011). Wahrnehmungsdimensionen von Stimme und Sprechweise. 22. Konferenz Elektronische Sprachsignalverarbeitung, Aachen, pp. 261–268.

Zuckermann, M., & Driver, R. E. (1989). What sounds beautiful is good: The vocal attractiveness stereotype. *Journal of Nonverbal Behaviour*, *13*, 67–82.

Chapter 2
Prosodic Aspects of the Attractive Voice

Andrew Rosenberg and Julia Hirschberg

Abstract A speaker's voice impacts listeners' perceptions of its owner, leading to inference of gender, age, personality, and even height and weight. In this chapter, we describe research into the qualities of speech that are deemed "attractive" by a listener. There are a number of ways that a person can be found attractive. We will review the research into what makes speakers attractive in the political and business domains, and what vocal properties lead to perceptions of trust. We then turn our attention to research into "likeability" and romantic attraction. While the lexical content of a speaker's speech is important to their attractiveness, we focus this survey on prosodic qualities, those acoustic properties that describe "how" the words are said rather than "what" the words are. Of course, attractiveness is subjective; what is attractive to one listener may not be to another. Properties of the listener and other contextual qualities can have a significant impact on the voices which are found to be attractive. The most comprehensive research in this topic includes analyses of both the speaker and the listener, since attraction is frequently a mutual phenomenon; when people are attracted to someone, they want to be found attractive in return. We will also summarize work that has investigated attraction dynamics in two-party conversations.

Keywords Likeability · Charisma · Political attractiveness · Business attractiveness · Romantic attraction · Speech prosody · Vocal attractiveness

A. Rosenberg (✉)
Google LLC, NYC, New York, NY, USA
e-mail: rosenberg@google.com

J. Hirschberg
Columbia University, NYC, New York, NY, USA
e-mail: julia@cs.columbia.edu

B. Weiss et al. (eds.), *Voice Attractiveness*, Prosody, Phonology and Phonetics,
https://doi.org/10.1007/978-981-15-6627-1_2