

RESEARCH REPORT

WHAT MAKES BUSINESS SPEAKERS SOUND CHARISMATIC? A CONTRASTIVE ACOUSTIC- MELODIC ANALYSIS OF STEVE JOBS AND MARK ZUCKERBERG



OPEN ACCESS

EDITED BY

- Miguel Oliveira, Jr. (UFAL)
- René Almeida (UFS)

REVIEWED BY

- Plínio Barbosa (UNICAMP)
- Tommaso Raso (UFMG)

ABOUT THE AUTHORS

- Oliver Niebuhr
Conceptualization, Formal Analysis,
Investigation, Methodology, Writing –
Original Draft, Writing – Review
& Editing.
- Alexander Brem
Conceptualization, Methodology.
- Jan Michalsky
Investigation, Writing – Original Draft,
Writing – Review & Editing.
- Jana Neitsch
Visualization, Validation, Writing –
Review & Editing.

DATES

- Received: 12/21/2019
- Accepted: 04/08/2020
- Published: 07/09/2020

HOW TO CITE

NIEBUHR, Oliver; BREM, Alexander;
MICHALSKY, Jan; NEITSCH, Jana
(2020). What Makes Business
Speakers Sound Charismatic?
A Contrastive Acoustic-Melodic
Analysis of Steve Jobs and Mark
Zuckerberg. *Cadernos de Linguística*,
v. 1, n. 1, p. 01-40.

Oliver NIEBUHR  

Centre for Industrial Electronics Mads Clausen Institute – University of
Southern Denmark

Alexander BREM  

Daimler Endowed Chaired Professor for Entrepreneurship in Technolo-
gy and Digitization – University of Stuttgart

Jan MICHALSKY  

Chair of Technology Management Friedrich-Alexander – Universität
Erlangen-Nürnberg

Jana NEITSCH  

Centre for Industrial Electronics Mads Clausen Institute – University of
Southern Denmark

ABSTRACT

Phonetic research on the prosodic sources of perceived charisma has taken a big step towards making a speaker's tone-of-voice a tangible, quantifiable, and trainable matter. However, the tone-of-voice includes a complex bundle of acoustic features, and a lot of parameters have not even been looked at so far. Moreover, all previous studies focused on political or religious leaders and left aside the large field of managers and CEOs in the world of business. These are the two research gaps addressed in the present study. An acoustic analysis of about 1,350 prosodic phrases from keynotes given by a more charismatic CEO (Steve Jobs) and a less charismatic CEO (Mark Zuckerberg) suggests that the

same tone-of-voice settings that make political or religious leaders sound more charismatic also work for business speakers. In addition, results point to further charisma-relevant acoustic parameters related to rhythm, emphasis, pausing, and voice quality - as well as to audience type as a significant context factor. The findings are discussed with respect to implications for future perception-oriented studies and perspectives for a computer-based measurement, assessment, and training of a charismatic tone of voice.

RESUMEN

La investigación sobre las características prosódicas de la percepción del carisma ha mostrado que el tono de voz de un orador es una característica tangible, cuantificable y entrenable. Sin embargo, el tono de voz incluye un conjunto complejo de rasgos acústicos y muchos parámetros no han sido estudiados hasta ahora. Además, los estudios previos se han centrado en el análisis del carisma de líderes políticos o religiosos y han dejado de lado el análisis de un gran número de managers y directores ejecutivos en el mundo de los negocios. En este estudio presentamos un análisis acústico de cerca de 1,350 frases prosódicas procedentes de discursos realizados por uno de los directores ejecutivos más carismáticos (Steve Jobs) y por uno menos carismático (Marc Zuckerberg). Los resultados sugieren que los ajustes del mismo tono de voz que hace que los líderes políticos y religiosos suenen más carismáticos también funcionen para oradores del mundo de los negocios. Además, los resultados muestran la relevancia de más parámetros acústicos, aparte del tono, para la percepción del carisma como son el ritmo, el énfasis, las pausas y la calidad de la voz - así como también el tipo de público como un factor significativo de contexto. Finalmente se discuten las implicaciones de los resultados de este estudio en relación a futuros estudios orientados a la percepción y con vista a la medición, evaluación y entrenamiento informáticos de un tono de voz carismático.

KEYWORDS

Persuasion; Charisma; Speech; Acoustic Analysis; Phonetics; Prosody; Steve Jobs; Mark Zuckerberg; English.

PALABRAS CLAVE

Persuasión; Carisma; Discurso; Análisis Acústico; Fonética; Prosodia; Steve Jobs; Mark Zuckerberg; Inglés.

INTRODUCTION

THE CONCEPT OF CHARISMA

Charisma is a complex and multifaceted phenomenon. Despite a rich and diverse research tradition it is still heavily surrounded by myths, ambiguities and misconceptions (ANTONAKIS *et al.*, 2016). The question about the nature of charisma can and should be tackled from different conceptual angles. One of them views charisma as an aspect of personality. Vergauwe *et al.* (2017) describe charisma as a specific personality-trait setting. Charismatic persons are stated to show extraverted personality facets like gregariousness or assertiveness, facets of openness to values or actions, conscientious facets such as achievement striving, and a lack of neurotic facets like anxiety, depression or self-consciousness. There are two major problems to this personality concept. Firstly, a charismatic personality may result in leader-like behavior. But this does not automatically include a charismatic style of communication (MICHALSKY *et al.*, 2020). Secondly, individuals who are ascribed a charismatic personality are often found to have narcissistic traits (ROGOZA; FATFOUTA, 2020) that counteract the effects of charisma, thus making it difficult to maintain a coherent and uniform concept of charismatic personality.

In an alternative line of thought charisma is described as attributional rather than a matter of personality. However, as is stressed by Antonakis *et al.* (2016) amongst others, this line of thought bears the risk of circular reasoning in that the description of charisma relies on its effects, which, in turn, substantiate its description. Thus, Antonakis *et al.* point out that charisma can be described neither by the attributes assigned to a speaker nor by the effects the speaker has on listeners. Instead, Antonakis *et al.* (2016, p. 304) describe charisma as an attribution to a specific communication style that is inherent to charismatic speakers and defined as “values-based, symbolic, and emotion-laden, leader” signals. A major caveat of this definition is the vague nature of the term ‘leader signals’. Although a certain vagueness is indeed required for the term to be used across various contexts, leader signalling can actually more substantially be defined with reference to three aspects that frequently occur in the literature (cf. WEBER, 1947; SHAMIR *et al.*, 1993, 1994; DEN HARTOG; VERBURG, 1997; CONGER *et al.*, 2000; EMRICH *et al.*, 2001; ANTONAKIS *et al.*, 2016; MICHALSKY; NIEBUHR, 2019). Firstly, charismatic leaders are characterized by passion, commitment and captivation or, in short, some sort of emotional involvement that is also transported by their speech. By conveying passion, commitment, and confidence, charismatic speakers transfer their emotional states and attitudes to their listeners via a process called emotional contagion (BONO; ILIES, 2006; BARSADÉ *et al.*, 2018). A second aspect is confidence or self-assurance. This is not to be confused with dominance or authority. The modern concept of charisma means to be persuasive without having power

and formal authority and, thus, without exerting hierarchical pressure on potential followers (cf. ANTONAKIS *et al.*, 2016). Rather, the strongest motivation to follow is created when listeners believe in the speaker and his/her ability to live up to the promises that are made (based on the values that are shared). This brings us to the third aspect of leadership signalling, *i.e.* competence. Charismatic leaders are competent in their field or, at least, they convey competence, for example, in the form of a structured and comprehensible delivery.

In summary, we arrive at a concept that defines charisma as a symbolic, emotion-laden and value-based communication style signaling leadership qualities such as commitment, confidence, and competence that affect followers' beliefs and behaviors in terms of motivation, inspiration, and trust.

THE RELEVANCE AND PHONETIC REALIZATION OF CHARISMA

Understanding the phenomenon of charisma is of “immense importance for society [...] because charismatic leaders wield enormous power and can use this for great good or evil” (ANTONAKIS *et al.*, 2016, p. 294). In addition to the fact that charisma is not a mysterious gift reserved for some chosen people (WEBER, 1947) but a learnable and improvable skill (TOWLER, 2003; ANTONAKIS; FENLEY; LIECHTI, 2011), it has been shown that being a charismatic speaker is relevant to many everyday situations. Being charismatic results in a more fruitful speed-dating or brainstorming output (PENTLAND; HEIBECK, 2010), gives students better learning performances (TOWLER, 2003), increases the chance of getting investors or raising start-up funding (DAVIES, 1954), makes a product or service more credible and likable to customers (GÉLINAS-CHEBAT; CHEBAT; VANINSKY, 1996), and functions as a professional-career catalyst (BODOW, 2002; JACQUART; ANTONAKIS, 2015).

Compared to what is known about who can be charismatic and what charismatic speakers can do, still relatively little is known about what charismatic speech exactly *is*. That is, how does charisma manifest itself in speech, or, put from a different angle, which aspects of the speech signal make a speaker sound more charismatic in the ears of the listener? Researchers with backgrounds in rhetoric, management, and (social) psychology have shed some light on this question in the recent past (*e.g.*, HOLLADAY; COOMBS, 1993; HOLLADAY *et al.*, 1998; CYPHERT, 2010; ANTONAKIS; FENLEY; LIECHTI, 2012; SØRENSEN, 2013). However, the descriptive labels that are often used in these contexts – such as “rich”, “animated”, “fluent”, and “powerful” – are hard to operationalize and replicate experimentally, and their instructive value for trainers and learners of charismatic speech is also limited (see NIEBUHR; TEGTMEIER; BREM, 2017). Digital speech-signal processing and analysis techniques can take a much more fine-grained approach that makes charisma in

speech a quantifiable and hence tangible research subject. For Antonakis *et al.* (2016: 308), the future of charisma research also lies in “unobtrusive and objective measures”. In a similar vein, but a few years earlier, Rosenberg and Hirschberg (2009) already called for an empirical definition of charismatic speech and, then, provided a first answer to their call by projecting ratings of perceived charisma attributes onto acoustic-prosodic features of speakers. This empirical foundation was further supported and enriched by similar studies of Biadys *et al.* (2008), Signorello *et al.* (2012, 2013), Niebuhr, Voße, and Brem (2016), Niebuhr and Fischer (2019), Niebuhr and Skarnitzl (2019), Niebuhr and Gonzalez (2019), Hiroyuki and Rathcke (2016), Novák-Tót, Niebuhr, and Chen (2017), Bosker (2017), Strangert and Gustafson (2008), D’Errico *et al.* (2013), Berger, Niebuhr and Peters (2017), and Jokisch *et al.* (2018).

These studies already identified a set of acoustic features that commonly distinguished more and less charismatic speakers. For example, charismatic speech was found to show an elevated rather than a lowered fundamental frequency (f_0) level as well as higher levels of vocal effort and intensity (TOUATI, 1993; STRANGERT; GUSTAFSON, 2008; BIADSY *et al.*, 2008; ROSENBERG; HIRSCHBERG, 2009; D’ERRICO *et al.*, 2013; BERGER; NIEBUHR; PETERS, 2017; JOKISCH *et al.*, 2018; NIEBUHR; SKARNITZL, 2019). In addition to higher parameter levels, there is also more variability in charismatic speech, for example, manifesting itself in a larger f_0 range (TOUATI, 1993; STRANGERT; GUSTAFSON, 2008; ROSENBERG; HIRSCHBERG, 2009; KREIMAN; SIDTIS, 2013) and a greater acoustic-energy dynamics (BOSKER, 2017). Further characteristics of charismatic speech are a higher speaking rate, shorter silent and fewer filled pauses (ROSENBERG; HIRSCHBERG, 2009; NIEBUHR; FISCHER, 2019). Lastly, more charismatic speakers partition their speech into smaller pieces of information by using more and shorter prosodic phrases (NIEBUHR; VOSSE; BREM, 2016) ideally with durations below that of the listeners auditory short-term memory (BADDELEY *et al.*, 2009).

Quantitative empirical insights like these have various possible applications in speech technology, ranging from persuasive and attractive text-to-speech synthesis (NIEBUHR; MICHALSKY, 2019; FISCHER *et al.*, 2019) to rhetorical training devices that detect, measure, and quantify a speaker’s charisma and then give detailed automatic feedback about which aspects of speech need to be improved for which purpose and how (NIEBUHR; TEGTMEIER; SCHWEISFURTH, 2019). Such a training device would be invaluable for entrepreneurs and various other business professionals like sales and call-center agents. Developing such a device on a solid experimental-phonetic basis is the authors’ goal.

The first basic obstacle that needed to be overcome on our way was that virtually all previous studies dealt with charismatic speakers from the fields of politics or religion, such as Jacques Chirac, Pope John Paul II, John Kerry, Arnold Schwarzenegger, and Donald

Trump (TOUATI, 1993; ROSENBERG; HIRSCHBERG, 2009; HIROYUKI; RATHCKE, 2016). Our target group, business figures, has almost entirely been disregarded so far.

One reason for this gap in business speakers could be that the term ‘charisma’ was used in religious contexts in the early 20th century (SOHM, 1923). Then, the seminal re-definition of the term by Weber (1947) made charisma a hidden innate personality trait that needs a societal crisis to emerge and then gives “leaders Salvationist qualities to deliver followers from great upheaval” (ANTONAKIS *et al.*, 2016, p. 295). This way of thinking included politicians in the group of potentially charismatic leaders, but it is still not applicable to business leaders. It is only since recently that Antonakis *et al.* (2016, p. 308) made charisma independent of any particular societal context or speaker group by defining it simply as a values-based emotion-laden leadership that is expressive in its transmission of information.

A further reason for the research gap in charismatic business speakers is the stronger mass media presence of political and religious figures and their inherent need to attract followers or voters through persuasive monologues in front of large audiences, see also Touati (1991) and Cyphert (2010). On this basis, it was more likely that religious and political figures would get associated with charisma in public opinion. However, this situation has changed as much as the definition of charisma. Nowadays, companies and their representatives have a huge financial influence and get more and more involved in political and social decision-making processes. Moreover, there is an increasing number of CEOs who no longer just represent their companies. Rather, they have become an integral part of the company’s brand image. In combination, these circumstances successively blur the distinction between politicians and CEOs.

QUESTION AND AIMS

Summarizing 1.2, it is time to address the question whether the prosodic characteristics of charismatic speech that the above cited researchers have found for political and religious leaders also apply to business leaders, *i.e.* CEOs. Based on this question, we pursue two aims with the present study.

- (1) We test by means of a prosodic analysis whether the multiparametric differences between a more and a less charismatic CEO are consistent with the known parameter changes that make political and religious leaders sound more charismatic.
- (2) We extend the analyzed set of prosodic parameters to aspects of rhythm, emphasis, pausing, and voice quality that have not been taken into account in previous studies.

Furthermore, we do not restrict the analysis to (semi-)automatically extractable features but include a number of prosodic features like emphatic-accent rate and type. Such temporally or spectrally more complex and phonology-related parameters also considerably affect a speaker's charismatic impact (NIEBUHR; THUMM; MICHALSKY, 2018) but require a great degree of manual annotation. Thus, they are missing in most previous analyses. By including both automatically and manually detectable prosodic features we arrive at a more comprehensible prosodic profile for investigating and eventually understanding and predicting speaker charisma. In this sense, the present between-speaker comparison also serves to identify prosodic features that need to be included in any further large-scale quantitative study on charismatic speech.

Addressing aims (1) and (2) will pave the way for subsequent studies and enable them to conduct a comprehensive series of perception experiments that not only quantify the effect size of prosodic parameter differences on perceived charisma. In addition, these experiments will also determine how powerful and sensitive each individual parameter is in triggering charisma.

THE TWO COMPARED SPEAKERS: STEVE JOBS AND MARK ZUCKERBERG

In order to pursue these aims, we selected two of the most well-known male US American CEOs of our time as our research objects, see Figure 1. One is Steve Jobs (SJ), who was famous for his charismatic speeches that were also scientifically studied in longitudinal and cross-sectional approaches (SØRENSEN, 2013; EMRICH, 2001). The other one is Mark Zuckerberg (MZ), whose speaking skills made Tobak (2012) even question the relevance of charisma in modern leadership. SJ and MZ are often named side by side as examples for differently charismatic business leaders, as in the CNN article of Sutter (2011) entitled "When it comes to presentation, Mark Zuckerberg is no Steve Jobs".

However, note that the two speakers do not constitute extreme poles on the charisma spectrum and were not intended to do so. We assume that MZ shows about average or slightly above average public-speaking skills. Thus, we are not comparing a charismatic speaker to an entirely uncharismatic speaker, but an exceptionally charismatic speaker (SJ) to an average reference speaker.



Figure 1. Results of the perception experiment on the charismatic impact of SJ and MZ.

Assessments of SJ's and MZ's charisma difference as they are paraphrased, for example, in Sutter's statement above were recently substantiated in a formal perception experiment by Mixdorff, Niebuhr and Hönemann (2018). The stimuli of this perception experiment were excerpts from the same keynote-speech material that is also analyzed in the present study. The stimuli were chosen to be free from strong attitudinal expressions, brand or product names, and other key words that could lead to the identification of SJ and MZ or their respective companies. Moreover, the stimuli were also comparable in content.

Nevertheless, we de-lexicalized the stimuli by low-pass filtering them at 600 Hz, in this way not only removing potential biases of verbal content but also of speaker identity. Low-pass filtering is a common method of speech de-lexicalization (MAREÛIL *et al.*, 2015). However, the 600 Hz low-pass filtering threshold was set relatively high and determined by the authors in a spiral-like trial-and-error progression. We aimed for a threshold that would be able to remove enough verbal content to preclude a spontaneous understanding of entire sentences and overarching meanings, while at the same time enabling participants to perceive basic charisma-related aspects of speech communication such as gender, expressiveness, emotions, style, prosodic phrasing, loudness (vocal effort), rhythm, stress, emphasis, etc. That is, preserving these aspects in the stimuli as far as possible was more important than consistently removing all phonemic traces. Individual words cannot have any charisma-relevant effect (ANTONAKIS *et al.*, 2016; ANTONAKIS; FENLEY; LIECHTI, 2011), and indirect charisma triggers like proper names or technical terms were not included in the stimuli. Therefore, a low pass filter at 600 Hz was better suited for our purposes than, for example, the PURR method, which is implemented as the "Sound-to-Hum" feature in PRAAT (SONNTAG; PORTELE, 1998). The PURR method reliably removes lexical content, but also gives the stimulus an artificial sound quality, in which most differences and changes in emotionality, expressiveness, style, etc. are lost.

A de-briefing questionnaire showed that only 5 (or 5.1 %) of the 98 participants, all of them proficient L2 speakers of English (*i.e.* at level C1), still recognized one or both of the speakers. These participants were removed from the dataset as familiarity with

the speaker was shown to affect charisma judgments of listeners (BIADSY, 2008). All remaining 93 participants also stated that they occasionally identified a few individual words in the stimuli, but that they are unable to understand what the speakers were talking about in general.

The results of the remaining 93 participants are summarized in Figure 2. The participants only had to do one short task that they performed on the basis of an online survey. The task was simply to listen to the de-lexicalized and anonymized speech stimuli and then rate, based on their auditory impression on a scale of 0-10, (i) how strong they think the speaker’s management skills are, (ii) how well the speaker performs as a leader, and (iii) how likely it is that they would dare to invest money in the speaker’s company. Ratings were made separately for each stimulus, but speaker order was randomized across stimuli and between participants. As is shown in Figure 2, SJ received higher ratings for all three aspects related to charismatic, persuasive speech. However, only the management and leadership differences were clearly pronounced and statistically significant at $p < 0.001$ according to t-tests for dependent samples ($t[92]=11.5/24.8$, $p < 0.001$).

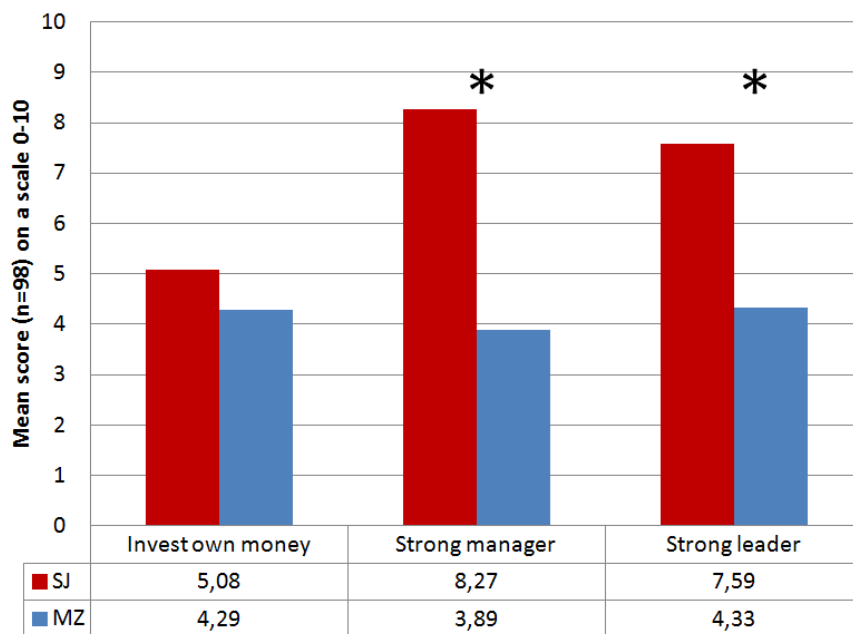


Figure 2. Pictures of SJ and MZ during their keynote speeches (edited and used under free CC license).

In the following, we describe the speech material that we used from each of the two speakers, the parameters that we analyzed, how we performed the analysis, and what results we got from the analysis. These results are then discussed in terms of our question and aims.

The significance of “melodic” (*i.e.* prosodic) features (BERNARD, 2012; HOTZ, 2014) is particularly stressed in interviews, observations, and anecdotes as well as in the scientific literature on charismatic speech. Therefore, prosody is also the focus of our study.

1. METHOD

1.1. SPEECH MATERIAL

Analogous to the study of Rosenberg and Hirschberg (2009), our analysis was based on sections of official and, thus, strongly conventionalized speeches. Moreover, taking into account Conger’s (1989) model of the complexity and contextual embedding of charisma (see also LEVINE; MUENCHEN; BROOKS, 2010), we further narrowed down the selection of speeches to a single genre. This ensured a constant speaking situation, content structure, and audience. The genre we used were product presentations, also because these globally broadcasted introductions of new products are particularly often referred to in the literature when it comes to speaker charisma (SØRENSEN, 2013).

For Steve Jobs, we used two of his most well-known and influential speeches: the presentation of the iPhone 4 in 2010 and the presentation of the iPad 2 in 2011. Each presentation included the following sections that occurred in the same order in both speeches:

- (1) Introduction: Welcoming. What was happening since the last presentation? What kind of problems arose with products and how have they been solved? What updates are available?
- (2) Main part I: Explanation of the company’s development and current market position as well as the success and significance of the previous product(s); advantages over competitors.
- (3) Main Part II: Presentation of the new product. Its main new features and innovations are demonstrated, their advantages for the user are emphasized, sometimes in comparison to competitors.
- (4) Main Part III: Presentation and demonstration of further related innovations (*e.g.*, apps); further information is provided on availability, price, and shipping of the presented product; accessories for the presented product are shown.
- (5) Summary and acknowledgments.

From sections (2) and (3) we extracted our speech data, as the speech in these middle sections was most consistent and free from effects of familiarization, boredom, and opening and closing addresses.

About 11 minutes of speech were extracted from each of the two sections, in approximately equal proportions from the iPhone 4 and the iPad 2 presentations (https://www.youtube.com/watch?v=z__jxoczNWc&t=500s; <https://www.youtube.com/watch?v=TGxEQhdi1AQ>); information on the exact amount of speech data is provided in the Additional Notes towards the end of this paper). The selection of speech data from sections (2) and (3) was random, but we disregarded parts of speech with insufficient audio quality, for example, due to music, applause, or noise. This gave us a total amount of 22 minutes of speech data, or about 12,000 individual speech sounds and 692 prosodic phrases, for our acoustic analyses. The extracted sound files had a sampling rate of 48 kHz and a 16-bit quantization and were saved on a computer in uncompressed WAV format (see also Additional Notes below).

Note that sections (2) and (3) addressed different audiences. Section (2) is relevant for investors, whereas section (3) showcases the product itself and is, thus, primarily oriented towards potential customers. This fact was taken into account in our study in order to detect (over and above our main aims, see 1.2) potential further differences between the customer-oriented and investor-oriented speech of SJ and MZ.

In the case of MZ, all speech samples were extracted from his keynotes at Facebook's "F8" events (cf. RUSLI, 2014). "F8" is Facebook's annual conference. It is meant to be a forum for highlighting milestones, advertising new features, and announcing the company's future plans and growth strategies. Accordingly, the "F8" keynotes given by MZ are structured similarly to those of more tangible (hardware) products, and, crucially, they also included customer-oriented and investor-oriented sections that met the same criteria as SJ's sections (2) and (3), the only difference being that MZ's investor-oriented sections target investors as well as app developers and entrepreneurs rather than just investors in the strict financial sense. That is, the customer-oriented samples of SJ and MZ shared, amongst other things, content structure and speaker intention. In this sense, they are as comparable as two natural spoken datasets can be. The same applies to the investor-oriented samples of SJ and MZ.

Samples representing customer-oriented and investor-oriented speech of Mark Zuckerberg were randomly selected from three separate presentations, again disregarding parts of speech with insufficient acoustic quality. Ten minutes of customer-oriented speech were selected from the key note presentation of 2011, and 5-6 minutes of each of Zuckerberg's F8 keynotes from 2014 and 2015 (see <https://www.youtube.com/watch?v=9r46UeXCzoU>; <https://www.youtube.com/watch?v=0oncilB-ZJA>; <https://www.youtube.com/watch?v=50x0JxhtEIA>). This gave us a total

speech sample of 21 minutes for acoustic analysis, consisting of about 13,700 speech sounds and 536 prosodic phrases. Information on the exact amount of speech data is provided in the Additional Notes towards the end of this paper. Like SJ's data, the extracted sound files of MZ were saved on a computer in the uncompressed WAV format with a sampling rate of 48 kHz and a 16-bit quantization.

1.2. ASSUMPTIONS

Based on prosodic analyses of political or religious speech, the studies of Touati (1993), Bidsy *et al.* (2008), D'Errico *et al.* (2013), Rosenberg and Hirschberg (2009), Signorello *et al.* (2012; 2013), among others, revealed a number of acoustic features that are correlated with perceived speaker charisma. In summary, higher f_0 levels and larger f_0 ranges, higher intensity levels and larger intensity variability, a higher speaking rate, shorter pauses, fewer disfluencies (filled pauses/hesitations and self-repairs), and shorter prosodic phrases are all positively correlated with a political speaker's perceived charisma.¹

Against this empirical background, the following assumptions are made and tested for our two business speakers SJ and MZ:

1. SJ's f_0 level and range are higher than those of MZ;
2. SJ's intensity level is higher and more variable than that of MZ;
3. SJ's speaking rate is higher than that of MZ;
4. SJ's pauses are shorter and the frequency of disfluencies is lower than for MZ;
5. SJ's prosodic-phrase durations are shorter than those of MZ.

In addition, given that charismatic speakers are said to have 'full' and 'durable' voices (MOREY, 2010) and are associated with expressive attributes like 'dynamic' and 'passionate' (SIGNORELLO *et al.*, 2012, 2013), the following additional parameters were included in our analysis:

¹ However, please see 4.1 for a critical discussion of the limits of such a uniform (perhaps even linear) relationship between changes in parameter levels and changes in perceived speaker charisma. For the sake of simplicity, and because it does indeed apply to the core ranges of prosodic parameters in speech communication, we take for granted in our assumptions 1-7 that parametric changes in a certain direction actually have a uniform effect on charisma.

- Four established voice-quality measures: jitter, harmonic amplitude difference (H1-H2, formant adjusted according to Iseli, Shue and Alwan [2007]), harmonics-to-noise ratio (HNR), and spectral emphasis, see Patel *et al.* (2011) for a discussion of these measures;
- Relative frequency (*i.e.* rate) of emphasized words, see Niebuhr (2010) for a discussion;
- Two robust and established rhythm metrics: %V and VarcoV, see Arvaniti (2012), and White and Mattys (2007), or Wiget (2010) for discussions of these measures.

These additional parameters led to the following additional assumptions:

6. SJ has a fuller and more durable voice than MZ which manifests itself in shorter pauses between prosodic phrases and, within prosodic phrases, lower jitter values but higher values of H1-H2, HNR, and spectral emphasis;
7. SJ has a more energetic and passionate way of speaking which manifests itself - beyond the measures (1)-(3) above - in a higher rate of emphasized words as well as in a higher rhythmic variability in terms of larger differences between minimum and maximum %V and VarcoV.

The term jitter describes the small period-to-period variation in f_0 and hence deviation of a speaker's voice from strict periodicity. The lower the jitter value the more harmonic and less trembling the voice is. H1-H2 represents the amplitude difference between f_0 and the first harmonic. Lower H1-H2 values are indicative of a pressed or strained voice. HNR quantifies clarity of a voice in terms of the energy ratio between the periodic components and the additive noise that a speaker's vocal-fold vibration generates. The higher the HNR value the more durable and clearer and less hoarse and rough a voice is. Spectral emphasis is an estimate of the spectral slope, *i.e.* the successive loss of acoustic energy across the ascending harmonics of a speaker's voice. Higher spectral-emphasis values mean a louder and fuller the voice (NIEBUHR; SKARNITZL, 2019; TRAUNMÜLLER; ERICKSSON, 2000).

Emphasized words are explained in more detail in 1.4. However, at this point it is clarified already that emphasis refers to phonetic (mainly prosodic) means of perceived prominence. Thus, emphasized words are those that stand out in perception more strongly than regular (nuclear) sentence-accented words, with the extra increase in perceptual salience being caused by some extra effort in phonatory and articulatory settings, timing, and dynamics (KÖHLER, 2006). That is, emphatic words are essentially a phenomenon of prosodic phonetics, unlike other non-phonetic means of emphasis like syntactic/verbal fronting (CORMACK; SMITH, 2000) that have not been included in the present analysis.

%V is the proportion of vowel segments within a prosodic phrase; VarcoV represents the variability (standard deviation) of vowel durations within a prosodic phrase, normalized in relation to the speaking rate of that phrase. With syllables being the rhythmical beats, more regular speech rhythms are characterized by higher %V and lower VarcoV values; more variable or irregular speech rhythms have lower %V and higher VarcoV values (see ARVANITI, 2012).

1.3. DATA ANNOTATION

In an initial step, the extracted customer-oriented and investor-oriented speech samples of each speaker were compiled and saved as separate WAV files. Each file represents a subsample.

Based on these subsamples, we carried out a manual annotation of the speech signal by means of the phonetic speech-processing tool PRAAT (BOERSMA, 2002). On an auditory basis, we first identified the prosodic phrases (*i.e.*, all coherent parts of speech in between two audible breaks, see the well-established “breath group” definition of Jones (1918) as well as modern concepts of phrasal annotation; JUN, 2005) in a subsample. Onsets and offsets were annotated, and, by means of these boundaries, the prosodic phrases were separated from other types of non-verbal signals, such as intended silent pauses, filled pauses, disfluent silent pauses, laughter, hesitational lengthening, and self-repairs. Intended silent pauses were represented by the label <p>. All other non-verbal vocalization signals counted as disfluencies and were represented by the label <hes>.

Note that filled pauses, laughter, and self-repairs are, of course, very different types of phenomena both respect to their communicative function and their implementation in speech interaction. Most phonetic annotation systems also keep them separate (HOUGH *et al.*, 2015). However, from the point of view of a keynote-speech monologue, they are similar insofar as they interrupt the constant, planned, and structured flow of information created by the speaker. It is only for this reason that we merged these different phenomena into a single category that we labeled ‘disfluencies’. Note further that the difference between intended silent pauses and disfluency phenomena was made on an auditory basis by a trained phonetician (the 1st author), also because “there are [currently] no reliable acoustic or articulatory indicators allowing one to distinguish between a fluent and disfluent pause” (MYERS, 2012: 4). Correspondingly, no fixed a priori criteria were defined to identify intended silent pauses. However, unlike all other pauses and disfluency phenomena, intended silent pauses identified by the annotator’s trained ear were typically those that occurred after chunks of speech that were grammatically well-formed both intonationally and lexically. Moreover, the onsets and offsets of intended si-

lent pauses were typically free from glottal stops or glottalization phenomena, and pause durations were >300 ms.

The prosodic phrases were annotated as interval units in PRAAT, and what the speaker said within each interval unit was associated with the phrase in the form of an orthographic transcription.

WebMAUS (see STRUNK; SCHIEL; SEIFART, 2014) was used to create the additional annotation level of individual sound segments. WebMAUS starts from the orthographic transcription, conducts an automatic grapheme-to-phoneme conversion and then assigns each phoneme to a portion of the speech signal in a forced-alignment procedure. As automatic segmentation tools like WebMAUS can be rather error-prone depending, for example, on the speaker's level of speech reduction, the outputs of WebMAUS were carefully manually checked and corrected, taking into account the guidelines that are formulated in the Principles of Phonetic Segmentation by Machač and Skarnitzl (2013). The manual checks were made in displayed time intervals of about 1.5 seconds so that the level of precision with which segment boundaries were set was 10 ms or smaller.

All annotations were saved in separate Textgrid files for the customer- and investor-related sound files for both speakers, see also Figure 3 and note 5 in section 6 for further information.



Figure 3. Example of the acoustic-analysis and annotation display. Shown is an excerpt of MZ's customer-oriented speech.

1.4. MEASUREMENTS

All acoustic measurements were conducted automatically in the speech signal by means of PRAAT scripts (see note 6 in section 6). The scripts were based on an analysis window of 40 milliseconds that was shifted in constant intervals of 10 milliseconds through the respective prosodic phrase. Thus, a new measurement was taken every 10 milliseconds until the end of a prosodic phrase was reached. A 40-millisecond window corresponds to the reliable default setting in PRAAT, *i.e.* a pitch-floor of 75 Hz. The window is long enough to

achieve a good compromise between measurement sensitivity and contour smoothing in the case of RMS intensity; and, for f_0 , it ensures always about 2-4 periods fall within a window so that reliable f_0 values can be determined by the autocorrelation algorithm.

The individual measurements were then converted into a single mean, standard deviation, and range per prosodic phrase. Duration measurements, including those for %V and VarcoV, were made with the accuracy of 1 millisecond and based on the annotated phrase and vowel boundaries in the speech signal. For all parameters, outliers were checked and, if necessary, corrected by manual measurements in the speech signal. In summary, we took the following measurements (i)-(x), tailored to address our assumptions.

- I. *Mean f_0 level* of a prosodic phrase (in Hz);
- II. *f_0 range* (in semitones, st), *i.e.* difference between the highest and lowest f_0 values within a prosodic phrase;
- III. *Intensity level* of a prosodic phrase (RMS, in dB);
- IV. *Intensity variation* of a prosodic phrase (in dB), in terms of the standard deviation of all individual measurements within the phrase;
- V. *Duration* of a prosodic phrase (in seconds, s);
- VI. *Duration* of an intended silent pause (in seconds, s);
- VII. *Duration* of a disfluency, *i.e.* filled pause, laughter, self-repair, etc. (in seconds, s);
- VIII. *Speaking rate* of a prosodic phrase (in syllables per second, syl/s);
- IX. *Voice quality* of a prosodic phrase in terms of jitter (RAP in %), spectral amplitude difference (H1-H2 in dB), harmonics-to-noise ratio (in dB), and spectral emphasis (in dB);
- X. *Speech rhythm* of a prosodic phrase in terms of the two vowel-based correlates %V and VarcoV; minimum and maximum values were taken per prosodic phrase. In the case of VarcoV, minimum and maximum were calculated by using the mean vowel duration $\pm \Delta V$ (rather than the mean vowel duration alone) as our normalization coefficients. In the case of %V, minimum and maximum are the values that result if all vowels had the same duration as the shortest or longest vowel within the prosodic phrase.

Since we compare two speakers of the same gender, it is important to note that normalizing measurements of mean f_0 level was not required. In fact, the very aim of these measurements was to make potential speaker-specific f_0 levels visible. In contrast, differences in the size of the f_0 range were to be measured independently of the speakers' f_0 levels. Obviously, a range of, for instance, 60 Hz is qualitatively not the same when meas-

ured at f0 levels of 120 Hz and 180 Hz. This is why we measured the mean f0 level in absolute Hz values, but the f0 range in relative semitone (st) values. Note further that the intensity level measurements were normalized for each speaker against a reference word. We used the word “so” to that end, because of its high frequency of occurrence, simple CV structure, and the fact that its overall energy level is representative of speech insofar as it combines the low acoustic energy a voiceless consonant with the high acoustic energy of a diphthongized vowel ([səʊ]). We calculated the average intensity across all instances of “so” and then divided the intensity level of each prosodic phrase by this reference value. We did this separately for all subsamples.

Regarding the tempo measurements, the concept of ‘speaking rate’ applied here is somewhere in between the established definitions of ‘articulation rate’ and ‘speaking rate’ (TSAO; WEISMER; IQBAL, 2006), *i.e.* syllables only (also net syllable rate) vs. syllables plus all intermitting pausing, breathing, and disfluency events. In our speaking-rate measures, longer pauses and intervals of in/exhalation are excluded as we measured the syllables per second within a prosodic phrase. On the other hand, we included shorter pauses as well as potential disfluencies in the phrase-internal tempo measurements. As prosodic phrase boundaries were determined on an auditory basis (see 2.3), there was no single numerical threshold for distinguishing between longer and shorter pauses. However, longer pauses were mainly >300 ms and shorter pauses <200 ms, in line with Heldner (2011). Measuring tempo in the strict sense of ‘articulation rate’ was not considered useful, firstly, because comparisons with previous studies (that virtually all measured speaking rate) would be hampered and, secondly, because our measure was to estimate to some degree the two speakers’ perceived tempo. Perceived tempo is better represented by speaking-rate than by articulation-rate measures, although none of them covers the full complexity of tempo perception in speech communication (KOREMAN, 2006).

Voice-quality measurements were taken at the center of the vowel segments. Unbiased comparisons of voice-quality measurements across subsamples are possible as open and closed as well as stressed and unstressed vowels occur with similar frequencies in each subsample (see HELDNER, 2011; TEIXEIRA; OLIVEIRA; LOPES, 2013; TEIXEIRA; FERNANDES, 2014).

In addition to the acoustic measurements (I)-(X), we also took two non-acoustic measurements (XI) and (XII). The latter two measurements were frequency counts made across all prosodic phrases. Disfluencies and emphatically highlighted words were identified on an auditory basis by a trained phonetician (the 1st author) who has 15 years of experience in research on melodic features.

XI. Relative frequency of *disfluencies* (in counts per minute, cpm);

XII. Relative frequency of *emphasized words* (in counts per minute, cpm).

Disfluencies include all those non-verbal phenomena that are not intended silent pauses, see 2.3 above. Based on the typology set up by Niebuhr (2010), emphasized words can take several phonetic forms. One frequent form is a strong lengthening of either the vowel of the accented syllable, as in “fantaaaaastic” or the consonant at the onset of the accented syllable, as in “rrrrreally grrrreat”. In addition, we also counted the doubling of words (“very very nice”) as instances of emphasis, as well as so-called accent chains, *i.e.* sequences of equally strong and often incrementally increasing or decreasing pitch accent peaks that are linked to words produced in a syllable-by-syllable fashion. An example of an accent chain is “Ev-ry-sin-gle-mo-ment”. Note that emphasized words contribute to the speech rhythm of a prosodic phrase, not least because emphasis manifests itself most clearly as a change of vowel duration (cf. NIEBUHR, 2010; KOHLER, 2006; KÜGLER, 2008; NIEBUHR *et al.*, 2012). Thus, we expect a certain co-variation between our two rhythm metrics %V and VarcoV and the frequency of emphasized words. Yet, emphasized words cannot simply be measured in terms of %V and VarcoV alone. It is important to count emphasized words separately in order to capture their contribution to a speaker’s perceived expressiveness.

Of course, identifying and counting disfluencies and emphasized words on an auditory basis is a subjective task that inevitably involves a certain rate of “false alarms” and “missed events” or, more technically, false positives and false negatives. Research on the annotation of emphatically stressed words shows that the resulting total error rate, in terms of inter-annotator agreement, is at about 10% for monologues, *i.e.* the type of speech to which also keynote speeches belong (see KÜGLER *et al.* 2015). For disfluencies, the error rate in terms of inter-annotator agreement is less than 5% (HOUGH *et al.* 2015). We used only one annotator, which supports a consistent annotation. The annotator was also trained phonetically and acted according to the requirement to annotate only unambiguous cases. Due to this conservative approach, the total error rate in our study should be clearly below 5% for both the disfluency and the emphatic-word count.

1.5. INFERENCE STATISTICS

We used unpaired t-tests for determining if the performances of SJ and MZ differ significantly along the individual acoustic measurements (*i-x*) and when addressing customers and investors. Taking into account the number of t-test comparisons, Bonferroni corrections of alpha-error levels were included. Moreover, *df*-levels were adjusted in the case of heterogeneous variances (determined by means of independent F-tests). Differ-

ences between the two frequency counts (xi) and (xii) were analyzed in terms of Z-score proportion tests.

Since we stated in 1.2 that more regular speech rhythms are associated with higher %V and lower VarcoV values, whereas the opposite applies to more irregular rhythms, we additionally calculated Pearson's correlation coefficients $\langle r \rangle$ between the two types of rhythm measurements for the customer- and investor-oriented subsamples of each speaker.

Note that for those measurements that were taken based on the unit of the prosodic phrase (*i.e.* all measurements except for the vi-vii and the frequency counts), the 689 prosodic phrases of SJ included 312 phrases in the customer-oriented and 377 phrases in the investor-oriented subsample. The total of 536 prosodic phrases produced by MZ constituted subsamples of 245 customer-oriented and 291 investor-oriented phrases.

As regards the intended silent pauses, there were $254+310 = 564$ tokens in the customer- and investor-oriented subsamples of SJ and $154+189 = 343$ tokens in the customer- and investor-oriented subsamples of MZ.

The number of phrases or other counted events is at the same time the samples size (n) in the corresponding statistical test. Given the conservative Bonferroni correction of p -values, the significance threshold was set to $p < 0.1$. Effect sizes are reported for all statistical tests.

2. RESULTS

2.1. BETWEEN-SPEAKER COMPARISONS: SJ VS. MZ

Figures 4 to 9 provide a graphical summary of the results. In terms of f_0 (see Figure 4), we found that SJ's mean f_0 levels were higher than those of MZ. This was true for both the customer-oriented sections ($_c$) – 214 Hz vs. 169 Hz – and the investor-oriented sections ($_i$) – 225 Hz vs. 187 Hz – of their speeches. Both differences are significant ($t_c[555]=7.92$, $p_c < 0.001$, $d_c=1.44$; $t_i[666]=12.55$, $p_i < 0.001$, $d_i=1.81$). Note that SJ's f_0 level is so high that it comes close to average f_0 levels that have been determined for female speakers (SYRDAL, 1996; PÉPIOT, 2013).

The differences in f_0 range between the two speakers are similarly clear and significant ($t_c[555]=8.04$, $p_c < 0.001$, $d_c=1.66$; $t_i[666]=10.36$, $p_i < 0.001$, $d_i=1.79$). The F_0 ranges of SJ's prosodic phrases were on average almost twice as large as those of MZ in both in the customer-oriented section – 24.9 st vs. 12.3 st – and the investor-oriented section – 21.2 st vs. 11.8 st.

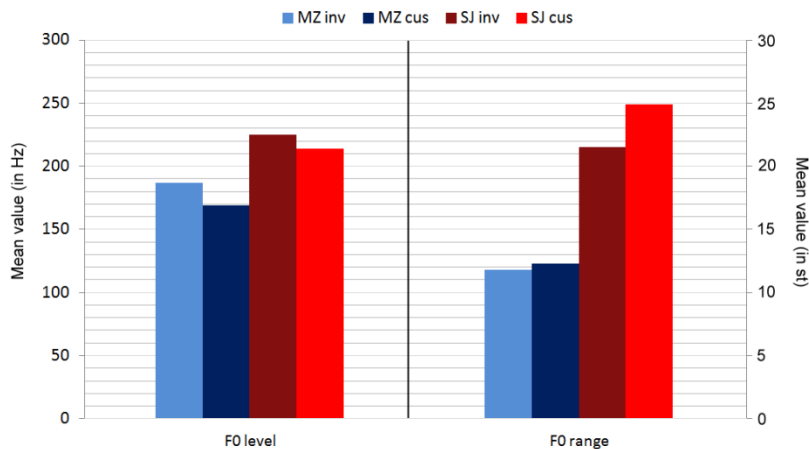


Figure 4. Results of the F0 analysis.

Regarding the intensity level (Figure 5), SJ's normalised mean intensity level (0.98) was found to be significantly higher ($t_c[555]=1.71$, $p_c<0.05$, $d_c=0.72$) than MZ's (0.95), but only in the customer-oriented subsamples. There was no significant intensity-level difference in the investor-oriented subsamples. Moreover, the two speakers did not differ with respect to intensity variation.

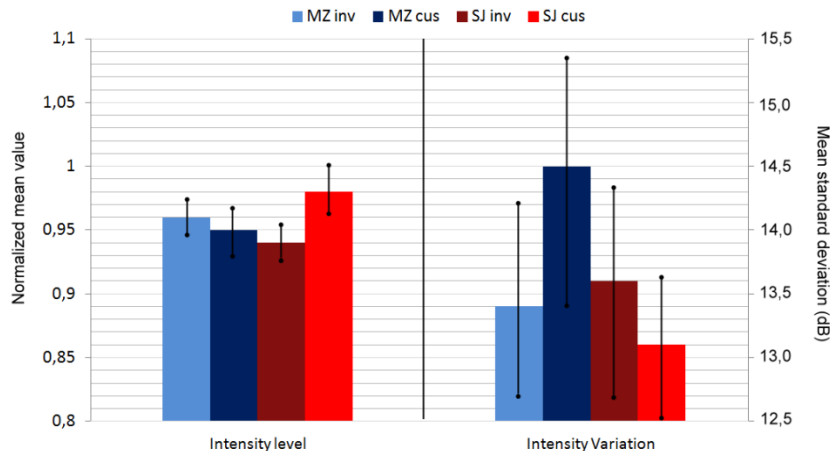


Figure 5. Results of the intensity analysis.

The duration measurements summarized in Figure 6 revealed SJ's prosodic phrases to be shorter than MZ's in the investor-oriented subsamples (1.39 s vs. 1.51 s) and even more so in the customer-oriented subsamples (1.25 s vs. 1.49 s). The duration differences in the customer- and investor-oriented subsamples are both significant, but with greater effect sizes in the customer-oriented ($t_c[555]=1.80$, $p_c<0.05$, $d_c=0.77$) than in the investor-oriented subsample ($t_i[666]=1.49$, $p_i<0.1$, $d_i=0.46$). Intended silent pauses were also shorter for SJ (0.56 s) than for MZ (0.83 s), but only in the customer-oriented subsamples

($t_c[410]=2.15$, $p_c<0.01$, $d_c=0.85$). In the investor-oriented subsamples, this difference was reversed. That is, here it were MZ's intended silent pauses that were shorter than those of SJ (0.62 s vs. 0.46 s: $t_i[499]=1.68$, $p_i<0.05$, $d_i=0.70$).

While SJ produced shorter prosodic phrases and, in one condition, also shorter intended silent pauses, MZ's speech was characterized by a considerably higher speaking rate. He exceeded 6 syl/s in the customer-oriented (6.16 syl/s) as well as in the investor-oriented speech (6.02 syl/s), whereas SJ was significantly slower (5.04 syl/s and 4.3 syl/s: $t_c[555]=13.51$, $p_c<0.001$, $d_c=1.93$; $t_i[666]=9.93$, $p_i<0.001$, $d_i=1.54$) in both subsamples. However, note that although SJ's speaking rate level was lower than that of MZ, it is still higher than mean rates found for "ordinary" speakers of American English from the phonetics literature (*e.g.* SYRDAL, 1996; ROBB; MACLAGAN; CHEN, 2004), particularly in the customer-oriented subsample.

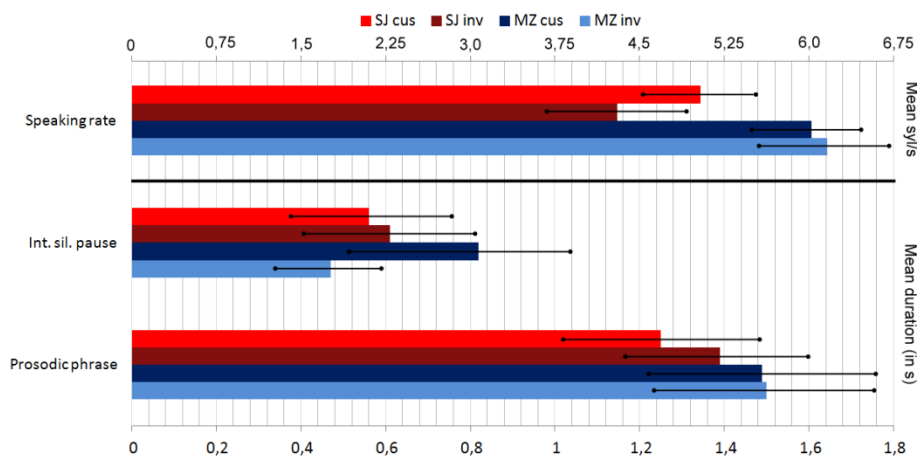


Figure 6. Results of the duration analysis.

The voice-quality differences between SJ and MZ (Figure 7) are clearly pronounced and involve all four acoustic parameters. We measured overall lower values of jitter (about 30 %) for SJ than for MZ, but this difference only reached statistical significance for the customer-oriented subsamples ($t_c[555]=2.79$, $p_c<0.01$, $d_c=1.11$). The differences in H1-H2 and HNR were also restricted to the customer subsample and showed on average at least 3 dB higher values for SJ than for MZ ($t_c[555]=9.96$, $p_c<0.001$, $d_c=1.62$; $t_c[555]=16.30$, $p_c<0.001$, $d_c=2.08$). Spectral emphasis was the only parameter that was significant for both subsamples, again with on average at least 3 dB (*i.e.* 100 %) higher mean values for SJ than for MZ ($t_c[555]=7.18$, $p_c<0.001$, $d_c=1.57$; $t_i[666]=5.44$, $p_i<0.001$, $d_i=0.96$).

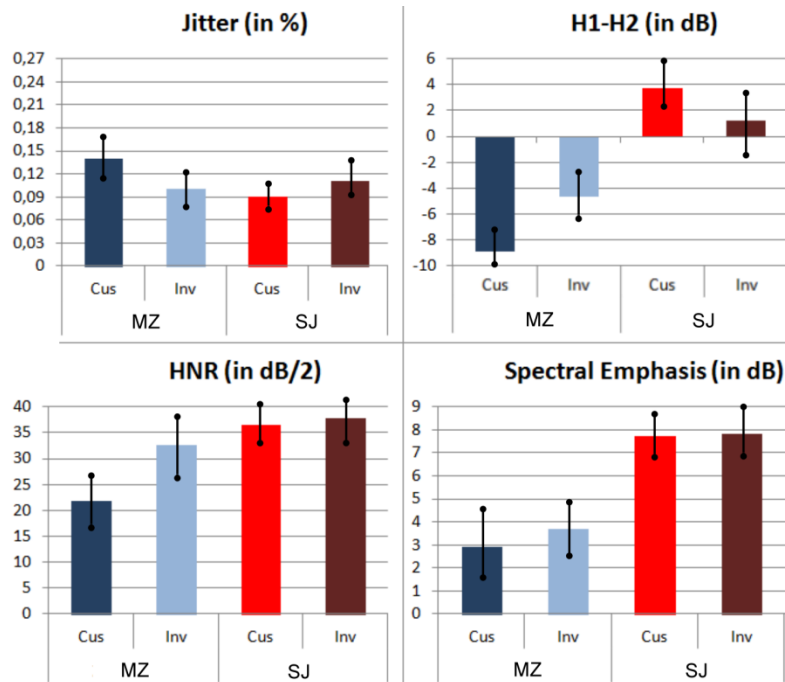


Figure 7. Results of the analysis of SJ's and MZ's voice-quality parameters.

The obtained rhythm measurements were statistically significantly different between SJ and MZ with respect to $\%V_{\min}$ ($t_c[555]=1.82$, $p_c<0.05$, $d_c=0.71$; $t_i[666]=1.76$, $p_i<0.05$, $d_i=0.63$), $\text{Varco}V_{\max}$ ($t_c[555]=1.88$, $p_c<0.05$, $d_c=0.88$; $t_i[666]=1.91$, $p_i<0.05$, $d_i=0.92$) and the customer-oriented measurements of $\%V_{\max}$ ($t_c[555]=2.47$, $p_c<0.01$, $d_c=1.24$), see Figure 8. SJ's values are a good half higher than those of MZ, *i.e.* 10 % on the $\%V_{\min}$ axis, 25 points on the $\text{Varco}V_{\max}$ axis, and in between 10-25 % on the $\%V_{\max}$ axis.

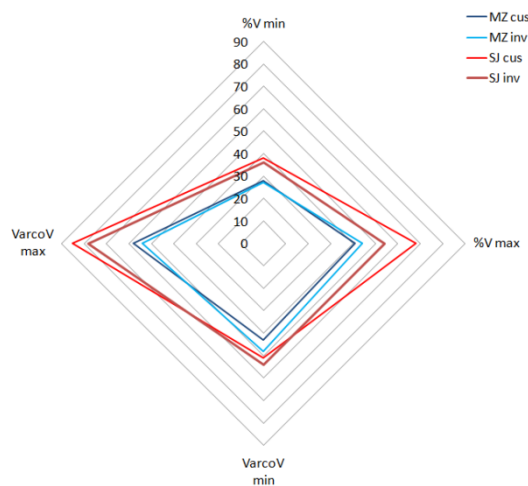


Figure 8. Results of the analysis of SJ's and MZ's rhythm parameters.

Furthermore, as our rhythm-related assumption 7 (see 2.2) referred to the differences between the maximum and minimum values of %V and VarcoV, *i.e.* to (two of) the side lengths of the quadrilateral in Figure 8, we ran four further t-tests for these max-min ranges. All four max-min ranges differed significantly between SJ and MZ in both the customer-oriented subsamples (%V_{max-min}: $t_c[555]=4.28$, $p_c<0.01$, $d_c=1.43$; VarcoV_{max-min}: $t_c[555]=3.99$, $p_c<0.01$, $d_c=1.51$) and the investor-oriented subsamples (%V_{max-min}: $t_i[666]=2.62$, $p_c<0.01$, $d_c=1.17$; VarcoV_{max-min}: $t_i[666]=1.75$, $p_c<0.05$, $d_c=1.05$).

Moreover, we tested for correlations between %V and VarcoV in the subsamples of each speaker (see 2.5) and found weak but significantly negative correlations between maximum VarcoV and minimum %V for both speakers (SJ: $r_c[310]=-0.17$, $p_c<0.001$, $d_c=0.35$; $r_i[375]=-0.22$, $p_i<0.001$, $d_i=0.45$; MZ: $r_c[243]=-0.15$, $p_c<0.001$, $d_c=0.30$; $r_i[289]=-0.13$, $p_i<0.001$, $d_i=0.26$). Likewise, minimum VarcoV and maximum %V were also weakly negatively correlated with each other (SJ: $r_c[310]=-0.21$, $p_c<0.001$, $d_c=0.43$; $r_i[375]=-0.16$, $p_i<0.001$, $d_i=0.32$; MZ: $r_c[243]=-0.24$, $p_c<0.001$, $d_c=0.49$; $r_i[289]=-0.17$, $p_i<0.001$, $d_i=0.35$). The correlation test between maximum VarcoV and maximum %V also came out significant (in terms of a positive correlation, SJ: $r_c[310]=0.24$, $p_c<0.001$, $d_c=0.49$; $r_i[375]=0.28$, $p_i<0.001$, $d_i=0.58$; MZ: $r_c[243]=0.19$, $p_c<0.001$, $d_c=0.39$; $r_i[289]=-0.20$, $p_i<0.001$, $d_i=0.41$), but not the one between minimum VarcoV and minimum %V.

The frequency counts (see Figure 9) showed that SJ's speech was much more lively in that he applied more than four times as many emphasis constructions than MZ in both the customer-oriented (70 vs. 16) and the investor-oriented (49 vs. 11) prosodic phrases ($Z\text{-score}_{c+i}$ across target audiences=12.38, $p_{c+i}<0.001$, $d_{c+i}=3.47$). As can be seen in Figure 9, this difference in the overall frequency of emphasized words relies to almost the same degree on all subtypes of emphasis. Proportionally, the difference is greatest for positive intensification, in which a word's accented vowel is lengthened underneath a high pitch-accent F0 plateau (as in "that's fantaaaaastic"). The proportional difference is smallest for the repetition of words ("very very important") and reinforcement, in which the pitch-accent f0 peak is steep and peaky and the onset consonant of the accented syllable is lengthened at the expense of the accented vowel (as in "that's rrrreally wwwonderful"). Reinforcement is the type of emphasis that both speakers used most often, followed by positive intensification in the case of SJ and accent chains ("Ev-ry-sin-gle-mo-ment") in the case of MZ.

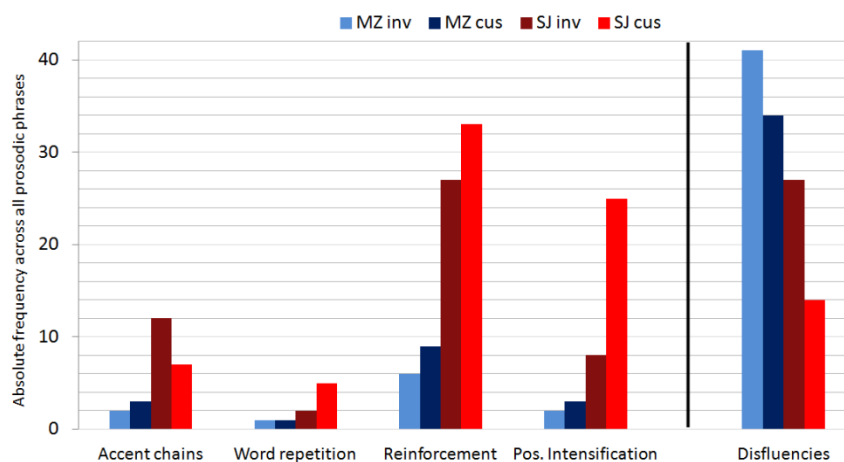


Figure 9. Results of the frequency of disfluencies and emphatic accents in the speech samples of SJ and MZ.

Finally, while SJ clearly outperformed MZ in terms of emphasized words, MZ takes the lead in the frequency of disfluencies (see Figure 9). We counted almost twice as many disfluency phenomena in MZ’s than in SJ’s speech, independently of whether the two speakers addressed customers (34 vs. 14) or investors (41 vs. 27; $Z\text{-score}_{c+i}$ across target audiences=3.55, $p_{c+i}<0.001$, $d_{c+i}=0.89$). Filled pauses like “ehm” and “err” were for both speakers clearly the most frequent type of disfluency. In addition, MZ also produced a notable number of self-repairs which were almost completely absent in the speech of SJ.

2.2. BETWEEN-AUDIENCE COMPARISONS: CUSTOMER- VS. INVESTOR-ORIENTED SPEECH

In 2.1, we reported those statistics that concerned the *between-speaker* comparisons *within* each subsample of customer-oriented and investor-oriented speech. Completing this results picture, Table 1 provides an overview of the *between-audience* comparisons *within* each speaker. That is, reported are those measurements that, according to a further series of t-tests, differed significantly between the customer-oriented and investor-oriented subsamples of each speaker. Two observations can be made in Table 1.

First, SJ used the charisma-associated prosodic features of his voice in a more context-specific way. That is, we found more significant differences between the customer- and investor-oriented sections in his speech than for MZ (the voice-quality parameters together represent one feature).

Second, if a parameter shows significant differences for both speakers, then these differences consistently go in opposite directions in the two subsamples. For example, SJ’s mean levels of F0, intensity, and H1-H2 as well as his rhythm metrics are all *higher* in the

customer-oriented than in the investor-oriented subsample. The exact opposite applies to MZ. His corresponding parameter levels are all significantly *lower* in the customer-oriented than in the investor-oriented subsample. Furthermore, SJ's phrase and intended silent pause durations both decrease from investor- to customer-oriented speech, whereas those of MZ increase.

	Steve Jobs	Mark Zuckerberg
	Difference of customer-oriented to investor-oriented speech	Difference of customer-oriented to investor-oriented speech
Pitch level	higher (p<0.001)	lower (p<0.01)
Pitch range	larger (p<0.01)	n.s.
Intensity level	higher (p<0.05)	lower (p<0.05)
Intensity range	n.s.	higher (p<0.01)
Phrase duration	shorter (p<0.001)	longer (p<0.05)
Speaking rate	faster (p<0.05)	n.s.
Disfluency dur.	n.s.	n.s.
Disfl. frequency	n.s.	lower (p<0.01)
Emph-acc frequ.	higher (p<0.001)	n.s.
Int. sil. pauses	shorter (p<0.05)	longer (p<0.001)
Jitter	n.s.	higher (p<0.05)
H1-H2	higher (p<0.001)	lower (p<0.001)
HNR	n.s.	lower (p<0.001)
Spectral Emph.	n.s.	lower (p<0.05)
%V	higher (p<0.001)	n.s.
VarcoV	higher (p<0.001)	n.s.

Table 1. Results summary of the t-tests and Z-score proportion tests used to compare for each speaker the measurements or frequency counts of the customer- and investor-oriented subsamples. The t-test's df values were between 152 and 688 for Steve Jobs and between 73 and 532 for Mark Zuckerberg.

3. DISCUSSION

3.1. AIM 1

Previous acoustic-prosodic research found higher f_0 levels and larger f_0 ranges, higher intensity levels and larger intensity variability, a higher speaking rate, fewer disfluencies (filled pauses/hesitations and self-repairs), and shorter prosodic phrases to be positively correlated with a political speaker's perceived charisma. Against this empirical background, the first and foremost aim of our study was to test by means of an acoustic-prosodic analysis whether the multiparametric differences between the two *business* speakers SJ and MZ (*i.e.* a less charismatic business speaker) are consistent with the above parameter changes and, thus, whether the patterns of charismatic political and religious charisma can be transferred, in a contrastive case study, to exemplary business speakers.

Disregarding non-significant findings, the evidence obtained from our contrastive case study suggests a positive answer to this question. SJ, who was popular for his charismatic speaking skills and whose (de-lexicalized and anonymized) speech stimuli were also associated with higher charisma-related performance ratings by listeners in a previous study, differed in multiple acoustic parameters from MZ, *i.e.* the less charismatic speaker according to public opinion and listener ratings. These differences match almost exactly with those described in the parametric summary above.

SJ spoke at a higher f_0 level and used a much larger f_0 range in his speech than MZ. SJ's intensity level was higher as well, at least in the customer-oriented section of his presentations. Compared to MZ, we also found shorter prosodic phrases and intended silent pauses, and fewer disfluencies in the analyzed speech samples of SJ.

There are only two aspects in which the differences between SJ and MZ do not match with the parametric summary above. The first one concerns speaking rate. MZ spoke significantly faster than SJ. At first glance this finding seems contradictory to the expected difference between SJ and MZ. On closer inspection, however, there is a plausible explanation for this finding. Even though SJ's speaking rate was on average lower than that of MZ, SJ's speaking rate is still higher than that of "ordinary" American English speakers (see SYRDAL, 1996; ROBB; MACLAGAN; CHEN, 2004). So, SJ is not a slow speaker either. The speaking rate of SJ, particularly in the customer-related parts of his keynotes, is still compatible with previous empirical evidence that charismatic speakers are faster than others. MZ, on the other hand, is not just a fast speaker, he is too fast a speaker. An average rate of 6.0 syllables per second or more approaches the magnitude that speakers produce when explicitly instructed to speak/read fast, for example, in a laboratory reading task (see DELLWO; WAGNER, 2003). In fact, MZ's speaking rate is so high that it caused a lot of

extreme segmental reductions. As is reported in Niebuhr and Gonzalez (2019) and Niebuhr (2020), MZ's vowel space is significantly smaller, *i.e.* his vowels are all more centralized and phonetically less distinct from one another than those of SJ. Moreover, MZ's speech contains 50 % more place assimilations of consonants and three times as many consonant lenitions as the speech of SJ. Such reductions make a speaker sound less educated and sincere, and more stressed, tired, and scatty (NIEBUHR, 2017).

In summary, our comparison was not between a slow and a fast speaker but between a fast and a very fast speaker. Rosenberg and Hirschberg (2009) have already pointed out that fast can be better than very fast. They assumed that correlations - positive and negative - between prosodic parameters and perceived speaker charisma are not linear, or they are only linear over a certain range of values. For parameter values outside this range, for which Niebuhr, Tegtmeier and Brem (2017) have introduced the term 'effectiveness window', the correlative relationship changes and can even reverse. We assume that this is true of MZ's speaking rate. Although he is above SJ's rate and the correlation between rate and charisma is positive, MZ's rate is so fast that the positive effect of increase speaking rate is reversed into a negative effect. Based on this assumption, there is no contradiction between the result of our contrastive case study and previous empirical evidence on the positive effect of an increased speaking rate.

A possible example of a parameter for which falling below a certain threshold reverses an otherwise positive charisma effect is the frequency of disfluencies. They occur for SJ about 300% less frequently than for other "ordinary" speakers in the phonetics literature (DUEZ, 1982; SHEINBERG, 2001), and they are also about half as frequent as those of MZ. Yet, SJ does still produce disfluencies, particularly filled pauses and hesitations; and this is probably a positive matter. Our ear is not used to process speech without any disfluencies. A certain minimal amount of disfluencies aids speech perception; and, moreover, studies have shown that a certain amount of disfluencies also makes a speaker sound addressee-oriented and friendly (SWERTS, 1998; FISCHER, 2000). We already have initial evidence for the fact that the complete absence of disfluencies makes a speaker sound less charismatic (NOVÁK-TÓT; NIEBUHR; CHEN, 2016; NIEBUHR; FISCHER, 2019).²

The second aspects in which the differences between SJ and MZ do not match with the parametric summary given at the beginning of 3.1 concerns the intensity variability.

2 Note that one of our reviewers asked us to discuss and specify in more detail, how the effectiveness windows are shaped for each acoustic-prosodic parameter of speaker charisma, and over and which value range they extend. Within our working group this is, in fact, ongoing research, in the course of which we have modeled the effectiveness windows for quite a few parameters already. We share the reviewer's opinion that "it would be very important" to specify some rough reference numbers for the shapes, ranges, and positions of these effectiveness windows in the paper. However, unfortunately we cannot comply with the reviewer's request in this regard, as the corresponding IP has been outsourced into a SDU spin-off company and is legally protected by this company.

We did not observe, as we had expected, a higher variability in SJ's than in MZ's intensity contours. This could mean that intensity variability is a less important feature in terms of charisma perception. Alternatively, taking into account that we consider MZ only *less* charismatic than SJ but not entirely *uncharismatic* (see also 3.3.2 below), our result could mean that, with intensity variability, we found one major source of MZ's charisma – the only source that is on a par with SJ's performance.

Overall, with respect to aim 1, we can conclude that our findings are in accord with the assumptions 1-5 that were put forward in 1.2. The prosodic profiles that have been worked out in previous studies and that make political or religious speakers sound more charismatic also apply to business speakers – at least to the two role models of business speakers that we analyzed here.

3.2. AIM 2

Inspired by rhetorical terms used to describe the tone-of-voice of charismatic (political) speakers, we expanded the parameters sets of previous studies to include parameters that measure the quality and durability of a speaker's voice as well as the passion and expressiveness conveyed by his/her speech. Given that SJ is a more charismatic speaker than MZ, the following additional assumptions were tested: Compared to MZ, SJ's speech would be characterized by lower jitter values but higher values of H1-H2, HNR, and spectral emphasis. In addition, we assumed to find more emphasized words in SJ's speech, as well as a higher rhythmic variability in terms of larger differences between minimum and maximum %V and VarcoV.

On the whole, these assumptions (6-7, see 1.2) were met by our acoustic analysis. SJ used more than four times as many emphasized words in his speech than MZ. Expressed in relative frequencies, SJ emphasized more than 5 words per minute, compared to less than 2 words per minute in the case of MZ. SJ's %V and VarcoV ranges were considerably larger than those of MZ. Since we took these measurements based on the unit of the prosodic phrase, what these results mean is that SJ's speech rhythm was more regular in one prosodic phrase and then became more irregular in the next prosodic phrase or vice versa. Yet, unlike we assumed, these results mainly rely on the maximum values of %V and VarcoV. That MZ's minimum values of %V and VarcoV were still lower or equally low as those of SJ can probably be explained by MZ's high speaking rate on the one hand and SJ's higher emphasis frequency on the other.

Note that the maximum values of %V and VarcoV that we found for MZ are even lower than many %V and VarcoV values of "ordinary" speakers reported in the phonetics literature (WHITE; MATTYS, 2007; RAMUS; NESPOR; MEHLER, 1999; PRIETO *et al.*, 2012). This

holds true although our calculations of %V and VarcoV in the present study yielded inherently higher values than in other acoustic studies on speech rhythm. Not least for this reason, MZ's small %V and VarcoV ranges, in combination with his low emphatic-word frequency, provide an explanation for why his speech sounds overall more monotonous and less variable and diverse in the authors' ears than the more animated, emotional, and passionate speech of SJ.

Our four voice-quality measures resulted in a consistent overall picture. As was expected, we found lower jitter values but higher values of H1-H2, HNR, and spectral emphasis for SJ as compared to MZ. That is, SJ's voice was not just louder, but also fuller, richer, and more harmonic and durable than that of MZ for whom the acoustic voice-quality measurements suggest a thinner, rougher, and more trembling voice, in line with our own perceptual impression. Also note that, although MZ talked in front of a big audience, all but one of his acoustic voice parameters (HNR) are even lower than values reported in the literature that were measured for people reading aloud or talking with each other in small silent rooms (HELDNER, 2003; TEIXEIRA; FERNANDES, 2014).

To conclude, we showed under aim 2 of the present study that it is necessary to expand the analysis beyond f0 and intensity to determine and evaluate a speaker's phonetic charisma profile. This fact further stresses that phonetic studies of charisma should not be done independently of previous rhetoric research. On the contrary, rhetorical terms and descriptions seem to be a valuable source inspiration for identifying what should be measured and how in a speaker's acoustic (or, more generally, phonetic) signal (see also NIEBUHR; TEGTMEIER; BREM, 2017).

3.3. DISCUSSION OF FURTHER ASPECTS

3.3.1. DIFFERENCES BETWEEN CUSTOMER-ORIENTED AND INVESTOR-ORIENTED SPEECH

Over and above pursuing our aims, it was additionally observed in the present study – for the first time as far as we are aware and in line with a modern, context-sensitive understanding of charisma (ANTONAKIS *et al.*, 2016; TOWLER, 2003) – that speakers produce quantitative changes in acoustic-melodic features of charisma depending on the audience they address. We found these changes for both SJ and MZ, but clearly more pronounced for SJ. The direction of these changes suggests that SJ's speech was more charismatic when addressing customers than when addressing investors. MZ's data show a tendency in the opposite direction, *i.e.* towards being more charismatic for investors than for customers.

One question in this context is who is the source of this difference, the speaker or the audience? If the audience is the source, then speakers only fulfill the expectations of different audience groups when they produce acoustic charisma differences between customers and investors. For example, does SJ actually sound equally charismatic towards investors and customers? His acoustic charisma signals are stronger towards customers, but what if customers also expect a stronger charisma signals than investors? Or is the speaker the source? That is, for example, is it SJ's (subconscious) intention to sound more charismatic towards customers than towards investors? We believe that the latter explanation is more plausible. First, if the audience were the source, then the differences between customers and investors would be the same, but SJ and MZ have realized different, even opposite, differences in their customer- and investor-oriented speech samples. Second, previous studies have shown that charisma and especially its acoustic cues are fairly consistent across genres, languages, and cultures (BIADSY *et al.*, 2008). So why should the addressed audience be an exception? Third, recall that, in MZ's case, investors were primarily developers/entrepreneurs, not lenders of capital. That is, they were basically programmers and hence part of M's own peer group. Similarly, SJ was known for having a very special and close relationship to his customers (cf. PAULMANN *et al.*, 2008). Thus, we suggest that SJ and MZ were more charismatic when they addressed that group of listeners they feel most comfortable in and/or more closely associated with.

If this explanation holds true and the speaker and his/her intention or feeling of belonging or identification are the source of audience-specific acoustic charisma differences, then this must be taken into account in any charisma training. Speakers would have to train to transfer the intrinsically more charismatic way of speaking from their most preferred and most favorite audience group to all other types of audience groups. We will check this in follow-up speech-production experiments with selected local entrepreneurs and different audience conditions.

3.3.2. ON THE CHARISMA OF MZ

The starting point of our study was that SJ is a more charismatic speaker than MZ according to both public opinion and experimental evidence. On this basis, we showed that this perceived charisma difference is systematically associated with acoustic-prosodic differences, which, in turn, match with differences that were found for political and religious speakers in previous phonetic studies. We also pointed out in this context, that MZ's parameter settings (*e.g.*, those of rhythm and voice quality) are partly even below levels that are produced by "ordinary" speakers that perform read speech or dialogue tasks in phonetic production studies.

Altogether, this could create the impression that MZ is an entirely uncharismatic speaker. Apart from the fact that we think of charisma as a quantifiable and thus gradual rather than a categorical (all-or-none) speaker characteristic, the performance of MZ deserves to be put into perspective. In particular, we want to stress again that MZ is on a par with SJ with respect to his intensity variability, a measure that many studies consider an important charisma factor. Further, we want to point out the fact that MZ actually also stands out against reference values of “ordinary” speakers from the phonetics literature in a number of his charisma-related prosodic features. Most notably, this concerns f0 level, prosodic-phrase duration, and the frequency of emphasized words. In his keynote speech samples, MZ’s f0 level is 40-60 Hz (*i.e.* 3-5 st) higher than the average f0 level of “ordinary” speakers engaged in reading tasks or dialogues (about 120 Hz, see SYRDAL, 1996; PAULMANN *et al.*, 2008; HAZAN; BAKER, 2011). MZ’s prosodic phrases are only about half as long as those of ordinary speakers (about 1.5 vs. 3.3 s; SWERTS; STRANGERT; HELDNER, 1996), while he produces at the same time twice as many emphatically highlighted words than “ordinary” speakers do in their reading or dialogue tasks (about 0.7 vs. 1.5 cpm; see PETERS, 2005; KOHLER, 2005; NIEBUHR *et al.*, 2015). Based on these facts, MZ is anything but an uncharismatic speaker. He is only less charismatic than SJ who is probably the role model of a charismatic speaker, at least among popular business speakers.

4. CONCLUSION AND OUTLOOK

Charisma is a key aspect of leadership and social interaction; and charismatic speech has been the subject of intensive research for centuries. However, what is still largely missing is a quantitative and objective line of research that, firstly, involves analyses of the acoustic signal that, secondly, focuses on business speeches such as product presentations, and that, thirdly, in doing so, advances the still fairly fragmentary evidence on the prosodic correlates of charismatic speech.

For the first time, to our knowledge, our analysis directly compared two influential business leaders of our time, SJ and MZ, in the way they gave major keynote speeches. Based on existing evidence that SJ is a more charismatic speaker than MZ (who is not completely uncharismatic either), our analysis primarily aimed at addressing the question whether SJ outperforms MZ in those acoustic dimensions that are known from previous studies to make political and religious speakers sound charismatic. The measurements we obtained from a total of 1,225 prosodic phrases of SJ and MZ provided a positive answer to that question. Thus, we concluded that those prosodic means that support the charismatic impression of a political or religious speaker also work in the fields of business and management. Of course, this is a far-reaching conclusion given

that we only analyzed two speakers for reasons pointed out in the introduction. But, firstly, we are currently running analyses on further business speakers both males and females and, so far, found no contradicting evidence that acoustic charisma patterns of business speakers differ from those of political and religious leaders; and, secondly, we consider this a reasonable generalization given how robust prosodic charisma features proved to be across cultures and political and religious camps. Yet, trying to identify nearly universal core features of charisma and separating them from features whose charisma interpretation is more variable across cultures, gender, age, leadership roles, product types, and presentation environments, will indeed a major supplementary task of follow-up studies (see BIADSY *et al.*, 2008).

Our second aim was to extend the scope of previous prosodic analyses to further potentially relevant charisma features. In this context, we provided initial production evidence that the frequency of emphasized words, rhythmic variables such as %V and VarcoV as well as voice-quality variables such as jitter, HNR, H1-H2, and spectral emphasis are related to a speaker's charismatic impression. Our future studies will further extend this set, including, for example, the amplitude modulation parameter of Bosker (2017). The order of magnitude with which our newly introduced parameters differed between SJ and MZ suggests that of some them (like emphasis and disfluency frequencies, rhythmic variability in terms of VarcoV, and the two voice measures H1-H2 and spectral emphasis) are perhaps even more important acoustic triggers of perceived charisma than established parameters like intensity level and phrase duration.

Assumptions like these illustrate how our aims 1 and 2 pave the way for a subsequent series of perception experiments that complement the acoustic analyses such that they provide perceptual evidence for the actual relevance of acoustic parameter differences in charisma perception, and, additionally, determine how powerful and sensitive each individual parameter (change) is in triggering charisma. The conclusions drawn with respect to aims 1-2 have brought us a lot closer to these experiments. We now have first supporting evidence that we can expect the same charisma profiles to work out for political, religious, *and* business speakers, and that rhythm, emphasis, and voice-quality should be integral parts of these profiles and hence also varied/controlled and tested in experimental stimuli.

Moreover, if we look at the consistency with which some parameters contribute to charisma perception across speakers and cultures (BIADSY *et al.*, 2008), if we look at how often some descriptive terms appear and are stressed in the rhetorical literature (SØRENSEN, 2013; GRIFFIN, 1992), and, finally, if we look at how large the parameter differences are that we found for SJ and MZ in the present study, then an overall fairly consistent picture emerges. This picture allows us to make initial assumptions as to which acoustic parameters are probably the most powerful and sensitive ones in charisma per-

ception. We assume that the following parameters belong to this group (no hierarchy is implied by the order of presentation): f0 level and range, speaking rate, the frequency of emphasized words and disfluencies, rhythmic variability (VarcoV in particular), intensity variability, and voice quality (in terms of H1-H2 and spectral emphasis). Weaker acoustic cues to charisma could hence be intensity level, as well as some voice-quality characteristics (jitter and HNR) and the durations of prosodic phrases, pauses, and disfluencies. These two short lists represent a promising starting point of subsequent perception experiments.

The types of parameters in the two lists suggest that perceived charisma is more a matter of articulation and phonation effort and its variability than of duration and timing. This nicely fits in with the general assessment of Hiroyuki and Rathcke (2016, p. 1) that “charisma [...] may be best understood as a skillful modulation of audio-visual prosody in social interaction”.

The further major focus of all our follow-up studies (next to extending our speaker set as well as empirical and parametric database) will be on these perception experiments. The initial experiments of Fischer *et al.* (2019) and Niebuhr and Michalsky (2019) in which the prosodic charisma profiles of talking robots were manipulated to test their effects on quantifiable behavioral variables of human listeners show how the perception of charisma features can be realized without using descriptive terms or obtrusive methods. In addition, we develop and evaluate methods for visualizing prosodic patterns such that they can correctly and consistently be produced by learners of charismatic speech (see NIEBUHR; NEITSCH, 2019) – because signal-based feedback on a speaker’s charismatic performance by a computer-based training device is not of much value without a properly functioning user interface. In combination with the growing body of evidence about which parameters play a role and how big these roles are, we get successively closer to the long-term goal, i.e. a computer-based training device (e.g., a smart phone app) that detects, measures, and quantifies a speaker’s charisma and then provides detailed automatic feedback about which aspects of speech a speaker needs to improve for which purpose and how.

A noteworthy additional finding of our study was that SJ and MZ produced significant differences between the customer-oriented and investor-oriented sections of their speeches. The results suggest that SJ’s speech was more charismatic when addressing customers than when addressing investors, and MZ’s data show a tendency in the opposite direction. This offers an interesting insight into the dynamics of charismatic speech, and it supports the assumption that charisma is not a constant speaker-specific characteristic but results from learnable and adjustable behavioral patterns. The latter wording finally implies that future studies of both production and perception also have to go beyond acoustic-prosodic features and include the whole domain of sound segment articulation, facial expressions, body movements and postures, and their interplay (*e.g.*, in terms of

timing and amplitude) with the speaker's verbal message. The pilot study of Hiroyuki and Rathcke (2016) represents a good starting point in this connection. In general, it is probably not exaggerated to conclude that, even almost 2,500 years after Aristotle's seminal works on pathos, ethos, and logos have laid the foundation for charisma research and training, what we know about charisma is still less than what we do not know. However, with modern computer-based and laboratory analysis, we have good chances to make a big step forward in this interdisciplinary question. Antonakis *et al.* (2016, p. 309) see "a very rosy future for the charisma construct", and we share their optimism.

5. ADDITIONAL NOTES

N1. The exact amount of speech data for Steve Jobs (taken from https://www.youtube.com/watch?v=z__jxoczNWc&t=500s; <https://www.youtube.com/watch?v=TGxEQhdi1AQ>)

- Investor-related section iPhone4: 4.437 min
- Investor-related section iPad2: 7.826 min
- Customer-related section iPhone4: 5.377 min
- Customer-related section iPad2: 4.220 min
- Speech data in total: 21.862 min

N2. The audio files of Steve Jobs can be downloaded here: 10.5281/zenodo.1187140

N3. The exact amount of speech data for Mark Zuckerberg (taken from the presentations in <https://www.youtube.com/watch?v=9r46UeXCzo>; <https://www.youtube.com/watch?v=0oncilB-ZJA>; <https://www.youtube.com/watch?v=50x0JxhtEIA>):

- F8 2011: 10.314 min
- F8 2014: 5.217 min
- F8 2015: 5.427 min
- Speech data in total: 20.958 min

N4. The audio files of Mark Zuckerberg can be downloaded here: 10.5281/zenodo.1187140

N5. The TextGrid files for the customer- and investor-related sound files for both speakers can be downloaded here: 10.5281/zenodo.1187140

N6. We would to thank Plinio Barbosa, Yi Xu, Eric Doty, Mietta Lennes, and Matthew Winn for publishing their scripts on the internet. Without this support, our investigation would not have been possible.

6. ACKNOWLEDGMENTS

The authors of this paper are greatly indebted to Jana Voße and Ester Novák-Tót for their selection and analysis of the presented speech material as well as to Stephanie Berger for her tremendous help in proof-reading the manuscript, formatting its text, and finalizing the reference section. Further special thanks are due to iö Valls and Pilar Prieto for their careful translation of the paper's abstract into Spanish. We would also like to thank our four anonymous reviewers for their insightful and constructive feedback on earlier drafts of this paper.

REFERENCES

- ANTONAKIS, John; BASTARDOZ, Nicolas; JACQUART, Philippe; SHAMIR, Boas. "Charisma: An ill-defined and ill-measured gift". *Annual Review of Organizational Psychology and Organizational Behavior*, v. 3, p. 293-319, 2016.
- ANTONAKIS, John; FENLEY, Marika; LIECHTI, Sue. "Can Charisma Be Taught? Tests of Two Interventions". *The Academy of Management Learning and Education*, v. 10, n. 3, p. 374-396, 2011.
- ANTONAKIS, John; FENLEY, Marika; LIECHTI, Sue. Learning charisma: Transform yourself into someone people want to follow. *Harvard business review*, v. 90, n. 6, p. 127-130, 2012.
- ARVANITI, Amalia. "The Usefulness of Metrics In the Quantification of Speech Rhythm". *Journal of Phonetics*, v. 40, n. 3, p. 351-373, 2012.
- BADDELEY, Alan D.; EYSENCK, Michael W.; ANDERSON, Mike. *Memory*. Hove, England: Psychology Press, 2009.
- BERGER, Stephanie; NIEBUHR, Oliver; PETERS, Benno. "Benno. Winning Over An Audience – a Perception-Based Analysis of Prosodic Features of Charismatic Speech". In: 43rd Annual Meeting of the German Acoustical Society, 2017, Germany: Kiel, p. 793-796.
- BERNARD, Andreas. Enthüllungsjournalismus. *Süddeutsche Zeitung Magazin* 40. 2012. Access: <http://sz-magazin.sueddeutsche.de/texte/anzeigen/38579/Enthuellungsjournalismus>. Cited 28 March 2020.
- BIADSY, Fadi; ROSENBERG, Andrew; CARLSON, Rolf; HIRSCHBERG, Julia B, STRANGERT, Eva. "A Cross-Cultural Comparison of American, Palestinian, and Swedish Perception of Charismatic Speech". In: 4th Conference on Speech Prosody, 2008, Brazil: Campinas, p. 579-582.
- BODOW, Steve. Charmed, I'm sure. USA Today. 09 December 2002. Available from: https://usatoday30.usatoday.com/money/jobcenter/workplace/successstrategies/2002-11-15-charismatic-leaders_x.htm. Access: 30 March 2020.
- BOERSMA, Paul. "Praat, a System for Doing Phonetics by Computer". *Glott international*, v. 5, n. 9, p. 341-345, 2002.
- BOSKER, Hans Rutger. The Role of Temporal Amplitude Modulations In The Political Arena: Hillary Clinton Vs. Donald Trump. In: 18th Annual Conference of the International Speech Communication Association (Interspeech), 2017, Sweden: Stockholm, p. 1-5.
- CABANE, Olivia Fox. *The Charisma Myth - How anyone can master the art and science of personal magnetism*. New York: Penguin, 2013.

CONGER, Jay A. *The charismatic leader: Behind the mystique of exceptional leadership*. San Francisco: Jossey-Bass; 1989.

CORMACK, Annabel; SMITH, Neil. "Fronting: The Syntax and Pragmatics of 'Focus' and 'Topic'". Technical Report, *UCL Working Papers in Linguistics*, p. 387-417. 2000.

CYPHERT, Dale. "The Rhetorical Analysis of Business Speech - Unresolved Questions". *Journal of Business Communication*, v. 47, n. 3, p. 346-368, 2010.

DAVIES, James C. "Charisma In the 1952 Campaign". *American Political Science Review*, v. 48, n. 4, p. 1083-1102, 1954.

DELLWO, Volker; WAGNER, Petra. "Relationships Between Rhythm and Speech Rate". Proc. 15th International Congress of Phonetic Sciences (ICPhS), 2003, Spain: Barcelona, p. 471-474.

DEN HARTOG, Deanne N.; VERBURG, Robert M. "Charisma and Rhetoric: Communicative Techniques of International Business Leaders". *The Leadership Quarterly*, v. 8, n. 4, p. 355-391, 1997.

D'ERRICO, Francesca; SIGNORELLO, Rosario; DEMOLIN, Didier; POGGI, Isabella. "The Perception of Charisma From Voice: a Cross-Cultural Study". In: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction IEEE (ACII), 2013, Can-ada: Toronto, p. 552-557.

DUEZ, Danielle. "Silent and Non-Silent Pauses In three Speech Styles". *Language and Speech*, v. 25, n. 1, p. 11-28, 1982.

EMRICH, Cynthia G; BROWER, Holly H; FELDMAN, Jack M; GARLAND, Howard. "Images In Words: Presidential Rhetoric, Charisma, and Greatness". *Administrative Science Quarterly*, v. 46, n. 3, p. 527-557, 2001.

FISCHER, Kerstin. *From Cognitive Semantics to Lexical Pragmatics: The Functional Polysemy of Discourse Particles*. Berlin: Mouton de Gruyter; 2000.

FISCHER, Kerstin; NIEBUHR, Oliver; JENSEN, Lars C.; BODENHAGEN, Leon. "Speech Melody Matters – How Robots Profit from Using Charismatic Speech". *ACM Transactions on Human-Robot Interaction (THRI)*, v. 9, n. 1, p. 1-29, 2019.

GANS, Joshua. "The Steve Jobs Theory of Customer Relations". *Harvard Business Review*. 13 December 2010. Access: <https://hbr.org/2010/12/customer-relations-from-the-to>. Cited 28 March 2020.

GÉLINAS-CHEBAT, Claire; CHEBAT, Jean-Charles; VANINSKY, Alexander. "Voice and Advertising: Effects of Intonation and Intensity of Voice on Source Credibility, Attitudes toward the Advertised Service and the Intend to Buy". *Perceptual and Motor Skills*, v. 83, n. 1, p. 243-262, 1996.

GRIFFIN Larry Don. "Walt Whitman's Voice". *Walt Whitman Quarterly Review*, v. 9, n. 3, p. 125-133, 1992.

HAZAN, Valerie; BAKER, Rachel. "Acoustic-Phonetic Characteristics of Speech Produced With Communicative Intent to Counter Adverse Listening Conditions". *Journal of the Acoustical Society of America*, v. 130, n. 4, p. 2139-2152, 2011.

HELDNER, Mattias. "On the Reliability of Overall Intensity and Spectral Emphasis as Acoustic Cor-Relates Of Focal Accents In Swedish". *Journal of Phonetics*, v. 31, n. 1, p. 39-62, 2003.

HELDNER, Mattias. "Detection Thresholds for Gaps, Overlaps and No-Gap-No-Overlaps". *Journal of the Acoustical Society of America*, v. 130, n. 1, p. 508-513, 2011.

HIROYUKI, Tanaka; RATHCKE, Tamara. "Then, What Is Charisma? The Role of Audio-Visual Prosody In L1 and L2 Political Speeches". In: *Phonetik & Phonologie im deutschsprachigen Raum*, 2016, Germany: Munich, p. 1-3.

HOLLADAY, Sherry J.; COOMBS, W. "Communicating Visions: an Exploration of the Role of Delivery In the Creation of Leader Charisma". *Management Communication Quarterly*, v. 6, n. 4, p. 405-427, 1993.

HOLLADAY, Sherry J.; COOMBS, W. "Speaking of Visions and Visions Being Spoken an Exploration of the Effects of Content and Delivery on Perceptions of Leader Charisma". *Management Communication Quarterly*, v. 8, n. 2, p. 165-189, 1994.

HOTZ, Robert Lee. "How to Train Your Voice to Be More Charismatic". *The Wall Street Journal*, 01 December 2014. Access: <https://www.wsj.com/articles/how-to-train-your-voice-to-be-more-charismatic-1417472214>. Cited 30 March 2020.

HOUGH, Julian; DE RUITER, Laura; BETZ, Simon; SCHLANGEN, David. "Disfluency and Laughter Annotation In a Light-Weight Dialogue Mark-Up Protocol". In: *The Disfluency in Spontaneous Speech (DISS 2015)*: ICPHS Satellite Meeting, 2015, Scotland: Edinburgh, p. 1-4.

ISELI, Markus; SHUE, Yen-Liang; ALWAN, Abeer. "Age, Sex, and Vowel Dependencies of Acoustic Measures Related to the Voice Source". *Journal of Acoustic Society of America*, v. 212, n. 2, p. 2283-2295, 2007.

JACQUART, Philippe; ANTONAKIS, John. "When Does Charisma Matter for Top-Level Leaders? Effect of Attributional Ambiguity". *Academy of Management Journal*, v. 58, n. 4, p. 1051-1074, 2015.

JOKISCH, Oliver; IAROSHENKO, Viktor; MARUSCHKE, Michael; DING, Hongwei. "Influence of Age, Gender and Sample Duration on the Charisma Assessment of German Speakers". In: *29th Conference on Electronic Speech Signal Process*, 2018, Germany: Ulm, p. 224-231.

JONES, Daniel. *An outline of English phonetics*, Cambridge: Cambridge University Press; 1918.

JUN, Sun-Ah. "Prosodic Typology". In: JUN, Sun-Ah. *Prosodic Typology*. Oxford: Oxford University Press, 2005. p. 430-458.

KOHLER, Klaus J. "Form and Function of Non-Pitch Accents". *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)*, v. 35a, p. 97-123, 2005.

KOHLER, Klaus J. "What Is Emphasis and How Is It Coded". In: *3rd International Conference of Speech Prosody*, 2006, Germany: Dresden, p. 748-751.

KOREMAN, Jacques. "Perceived Speech Rate: the Effects of Articulation Rate and Speaking Style In Spontaneous Speech". *Journal of Acoustical Society of America*, v. 119, n. 1, p. 582-296, 2006.

KÜGLER, Frank. "The Role of Duration as a Phonetic Correlate of Focus". In: *4th International Conference of Speech Prosody*, 2008, Brazil: Campinas, p. 591-594.

KÜGLER, Frank; SMOLIBOCKI, Bernadett; ARNOLD, Denis; BAUMANN, Stefan; BRAUN, Bettina; GRICE, Martine; JANNEDY, Stefanie; MICHALSKY, Jan; NIEBUHR, Oliver; PETERS, Jörg; RITTER, Simon; RÖHR, Christine T.; SCHWEITZER, Antje; SCHWEITZER, Katrin; WAGNER, Petra. DIMA: "DIMA: Annotation Guidelines for German Intonation". In: *18th International Congress of Phonetic Sciences*, 2015, Scotland: Edinburgh, p. 317-320.

LEVINE, Kenneth J; MUENCHEN, Robert A; BROOKS, Abby M. "Measuring Transformational and Charismatic Leadership: Why Isn't Charisma Measured?". *Communication Monographs*, v. 77, n. 4, p. 576-591, 2010.

MACHAČ, Pavel; SKARNITZL, Radek. *Principles of Phonetic Segmentation*. Praha: Epocha Publishing House; 2013.

MAREÜIL, Philippe Boula de; RILLIARD, Albert; LEHKA-LEMARCHAND, Iryna; MAIRANO, Paolo., & LAI, Jean-Pierre. "Falling Yes/No Questions In Corsican French and Corsican: Evidence for a Prosodic Transfer". In: *Prosody and Language in Contact*. Berlin-Heidelberg: Springer, 2015, p. 101-122.

MIXDORFF, Hansjörg; NIEBUHR, Oliver; HÖNEMANN, Angelika. "Model-Based Prosodic Analysis of Charismatic Speech". In: *9th International Conference of Speech Prosody*, 2018, Poland: Poznan, p. 1-5.

MOREY, Sean W. *Delivery @ Machines: Toward a Rhetoric and Decomposition of New Media*, PhD Thesis. USA: University of Florida, 2010.

MORTENSEN, Kurt W. *The laws of charisma: How to captivate, inspire, and influence for maximum success*. New York: Amacom Books, 2011.

MYERS, Julia. "The Behavior of Hand and Facial Gestures at Pauses In Speech". Bachelor Thesis, Yale University, New Haven, CT, USA. 2012. Access: <https://ling.yale.edu/sites/default/files/files/alumni%20senior%20essays/Myers%20Senior%20Thesis%20Final.pdf>. Cited 28 March 2020.

NIEBUHR, Oliver. "On the Phonetics of Intensifying Emphasis In German". *Phonetica*, v. 67, n. 3, p. 170-198, 2010.

NIEBUHR, Oliver. "Clear Speech - Mere Speech? How Segmental and Prosodic Speech Reduction Shape the Impression that Speakers Create on Listeners". *In: 18th Annual Conference of the International Speech Communication Association (Interspeech)*, 2017, Sweden: Stock-holm, p. 894-898.

NIEBUHR, Oliver. "'Space Fighters' on Stage-How F1 and F2 Vowel-Space Dimensions Contribute to Perceived Speaker Charisma". *In: SIEGERT, Ingo; BOCK, Ronald; WENDEMUTH, Andreas. Studentexte zur Sprachkommunikation*. TUDPress. 2020, p. 265-277.

NIEBUHR, Oliver; FISCHER, Kerstin. "Do Not Hesitate! – Unless You Do It Shortly Or Nasally: How the Phonetics of Filled Pauses Determine their Subjective Frequency and Perceived Speaker Performance". *In: 20th Annual Conference of the International Speech Communication Association (Interspeech)*, 2019, Austria: Graz, p. 544-548.

NIEBUHR, Oliver; GONZALEZ, Simon. "Do Sound Segments Contribute to Sounding Charismatic? Evidence from Acoustic Vowel Space Analyses of Steve Jobs and Mark Zuckerberg". *International Journal of Acoustics and Vibration*, v. 24, n. 2, p. 343-355, 2019.

NIEBUHR, Oliver; JARZABKOWSKA, Patrycja; LORENZ, Uta; SCHULZ, Christoph; SODI-GOV, Farhod. "Say It Again, Sam! The Prosodic Profiles of Emphatic Reduplication In German". *In: 6th International Conference on Speech Prosody*, 2012, China: Shanghai, p. 258-261.

NIEBUHR, Oliver; MICHALSKY, Jan. "Computer-Generated Speaker Charisma and Its Effects on Human Actions in a Car-Navigation System Experiment - or How Steve Jobs' Tone of Voice Can Take You Anywhere". *In: MISRA, Sanjay; TORRE, Carmelo; TARANTINO, Eufemia; APDUHAN, Bernady O.; GERVASI, Osvaldo; MURGANTE, Beniamino; STANKOVA, Elena; KORKHOV, Vladimir; ROCHA, Ana Maria A. C.; TANIAR, Da-vid. Lecture Notes in Computer Science*. New York: Springer Nature. 2019, p. 375-390.

NIEBUHR, Oliver; NEITSCH, Jana. "Raise Your Skills, Then Raise Your Voice Comparing Speech-Melody Visualization Tools For Acoustic Leadership Training". *In: 2nd Conference on Teaching for Active Learning*, 2019, Denmark: Odense, 1-2.

NIEBUHR, Oliver; PETERS, Benno; LANDGRAF, Rabea; SCHMIDT, Gerhard. "The Kiel Cor-Pora of 'Speech & Emotion' – a Summary". *In: 41st conference of the German acoustical society (DAGA)*, 2015, Germany: Nuremberg, p. 1-4.

NIEBUHR, Oliver; SKARNITZL, Radek. "Measuring a Speaker's Acoustic Correlates of Pitch-But Which? A Contrastive Analysis Based on Perceived Speaker Charisma". *In: 19th International Congress of Phonetic Sciences*, 2019, Australia: Melbourne, p. 1774-1778.

NIEBUHR, Oliver; TEGTMEIER, Silke; BREM, Alexander. "Advancing Research and Practice In Entrepreneurship Through Speech Analysis – From Descriptive Rhetorical Terms to Phonetically Informed Acoustic Charisma Metrics". *Journal of Speech Sciences*, v. 66, n. 1, p. 3-26, 2017.

NIEBUHR, Oliver; THUMM, Jana; MICHALSKY, Jan. "Shapes and Timing In Charismatic Speech - Evidence From Sounds and Melodies". *In: 9th International Conference of Speech Prosody*. 2018, Poland: Poznan, p. 1-5.

NIEBUHR, Oliver; VOSSE, Jana; BREM, Alexander. "What Makes a Charismatic Speaker? A Computer-Based Acoustic-Prosodic Analysis of Steve Jobs Tone of Voice". *Computers in Human Behavior*, v. 64, p. 366-382, 2016.

NOVÁK-TÓT, Eszter; NIEBUHR, Oliver; CHEN, Aoju. "A Gender Bias In the Acoustic-Melodic Features of Charismatic Speech?". *In: 18th Annual Conference of the International Speech Communication Association (Interspeech)*, 2017, Sweden: Stockholm, p. 2248-2252.

PATEL, Sonja; SCHERER, Klaus R; BJÖRKNER, Eva; SUNDBERG, Johan. "Mapping Emotions Into Acoustic Space: the Role of Voice Production". *Biological Psychology*, v. 87, n. 1, p. 93-98, 2011.

PAULMANN, Silke; SCHMIDT, Patricia; PELL, Marc; KOTZ, Sonja A. "Rapid Processing of Emotional and Voice Information as Evidenced by ERPs". *In: 4th International Conference of Speech Prosody*, 2008, Brazil: Campinas, p. 205-209.

PENTLAND A, HEIBECK T. "Honest Signals – How They Shape our World". Cambridge: MIT Press; 2010.

PÉPIOT, Erwan. "Voice, Speech and Gender: Male-Female Acoustic Differences and Cross-Language Variation In English And French Speakers". *In: 15th Rencontres Jeunes Chercheurs de l'ED 268*, 2013, France: Paris, p. 1-13.

PETERS, Benno. "The Database – The Kiel Corpus of Spontaneous Speech". *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)*, v. 35a, 1-6, 2005.

PRIETO, Pilar; DEL MAR VANRELL, Maria; ASTRUC, Lluïsa; PAYNE, Elinor; POST, Brechtje. "Phonotactic and Phrasal Properties of Speech Rhythm. Evidence from Catalan, English, and Spanish". *Speech Communication*, v. 54, n. 6, p. 681-702, 2012.

RAMUS, Frank; NESPOR, Marina; MEHLER, Jacques. "Correlates of Linguistic Rhythm In The Speech Signal". *Cognition*, v. 73, n. 3, p. 265-292, 1999.

ROBB, Michael P; MACLAGAN, Margaret A; CHEN, Yang. "Speaking Rates of American and New Zealand Varieties of English". *Clinical Linguistics & Phonetics*, v. 18, n. 1, p. 1-15, 2004.

ROSENBERG, Andrew; HIRSCHBERG, Julia. "Charisma Perception from Text and Speech". *Speech Communication*, v. 51, n. 7, p. 640-655, 2009.

RUSLI EM. "Facebook to Hold Developer Conference on April 30". *The Wall Street Journal*. 08 March 2014. Access: <http://blogs.wsj.com/digits/2014/03/08/facebook-to-hold-f8-developer-conference-on-april-30>. Cited 30 Mach 2020.

SHEINBERG, Elizabeth. "To 'Errrr' Is Human: Ecology and Acoustics of Speech Disfluencies". *Journal of the International Phonetic Association*, v. 31, n. 1, p. 153-169, 2001.

SIGNORELLO, Rosario; D'ERRICO, Francesca; POGGI, Isabella; DEMOLIN, Didier. How Charisma Is Perceived from Speech: A Multidimensional Approach. Privacy, Security, Risk and Trust (PASSAT). *In: International Conference on Social Computing (Social-Com)*, 2012, The Netherlands: Amsterdam, p. 435-440.

SIGNORELLO, Rosario; D'ERRICO, Francesca; POGGI, Isabella; DEMOLIN, Didier; MAIRANO, Paolo. "Charisma Perception In Political Speech: a Case Study". *In: International Conference on Speech and Corpora (GSCP 2012)*, 2013, Florence, Italy, p. 343-348.

SOHM, R. *Kirchenrecht* [Church Law]. Berlin: Duncker & Humblot; 1923.

SONNTAG, Gerit P; PORTELE, Thomas. "PURR - A Method For Prosody Evaluation and Investigation". *Computer Speech & Language*, v. 12, n. 4, p. 437-451, 1998.

SØRENSEN, Lotte Skjøttgaard. "How to Grow an Apple: Did Steve Jobs Speak Apple to Success?". Master Thesis, University of Aalborg, Denmark. 2013.

STRANGERT, Eva; GUSTAFSON, Joakim. "What Makes a Good Speaker? Subject Ratings, Acoustic Measurements and Perceptual Evaluations". *In: 9th Annual Conference of the International Speech Communication Association (Interspeech)*, 2008, Australia: Brisbane, p. 1688-1691.

STRUNK, Jan; SCHIEL, Florian; SEIFART, Frank. "Untrained Forced Alignment of Transcriptions and Audio for Language Documentation Corpora using WebMAUS". *In: 9th International Conference on Language Resources and Evaluation (LREC'14)*, 2014, France: Paris, p. 3940-3947.

SUTTER, John D. "When It Comes to Presentation Mark Zuckerberg Is No Steve Jobs". CNN. 08 July 2011. Access: <http://edition.cnn.com/2011/TECH/innovation/07/07/zuckerberg.facebook.presentation/index.html>. Cited 28 Mach 2020.

SWERTS, Marc. "Filled Pauses as Markers of Discourse Structure". *Journal of Pragmatics*, v. 30, n. 4, p. 485-496, 1998.

SWERTS, Marc; STRANGERT, Eva; HELDNER, Mattias. "F0 Declination In Read-aloud and Spontaneous Speech". *In: 6th International Congress of Spoken Language Processing*, 1996, USA: Philadelphia, p. 1501-1504.

SYRDAL, Ann K. "Acoustic Variability In Spontaneous Conversational Speech of American English Talkers". *In: 4th International Conference on Spoken Language Processing*, 1996, USA: Philadelphia, p. 438-441.

TEIXEIRA, João P; FERNANDES, Paula O. Jitter. "Shimmer and HNR Classification Within Gender, Tones and Vowels In Healthy Voices". *Procedia Technology*, v. 16, p. 1228-1237, 2014.

TEIXEIRA, João P; OLIVEIRA, Carla; LOPES, Carla. "Vocal Acoustic Analysis – Jitter, Shimmer and HNR Parameters". *Procedia Technology*, v. 9, p. 1112-1122, 2013.

TOBAK, Steve. "Charisma: An Obsolete Leadership Quality". *CBS News: Moneywatch*. 01 May 2012. Access: <http://www.cbsnews.com/news/charisma-an-obsolete-leadership-quality>. Cited 30 March 2020.

TOUATI, Paul. "Prosodic Aspects of Political Rhetoric". In: ESCA Workshop on Prosody, 1993. Sweden: Lund, p. 168–171.

TOUATI, Paul. "Temporal Profiles and Tonal Configurations In French Political Speech. Lund Working Papers In Linguistics". Lund University, Department of Linguistics, v. 38, 205-2019, 1991.

TRAUNMÜLLER, Hartmut; ERIKSSON, Anders. "Acoustic Effects of Variation In Vocal Effort by Men, Women, And Children". *Journal of the Acoustical Society of America*, v. 107, n. 6, p. 3438-3451, 2000.

TSAO, Ying-Chiao; WEISMER, Gary; IQBAL, Kamran. "Interspeaker Variation In Habitual Speaking Rate: Additional Evidence". *Journal of Speech, Language, and Hearing Research*, v. 40, n. 4, p. 1156-1164, 2006.

TOWLER, Annette J. "Effects of Charismatic Influence Training on Attitudes, Behavior, and Performance". *Personnel Psychology*, v. 56, n. 2, p. 363–381, 2003.

WEBER, Max. *The Theory of Social and Economic Organization*, New York: The Free Press of Glencoe, 1947.

WHITE, Laurence; MATTYS, Sven L. "Calibrating Rhythm: First Language And Second Language Studies". *Journal of Phonetics*, v. 35, n 4, p. 501-522, 2007.

WIGET, Lukas; WHITE, Laurence; SCHUPPLER, Barbara; GRENON, Isabelle; RAUCH, Olesya; MATTYS, Sven L. "How Stable are Acoustic Metrics of Contrastive Speech Rhythm?". *Journal of the Acoustical Society of America*, v. 127, n. 3, p. 1559-1569, 2010.