“High on Emotion“? How Audio Codecs Interfere with the Perceived Charisma and Emotional States of Men and Women

Oliver Niebuhr[1], *Ingo Siegert*[2]

[1] *Centre for Industrial Electronics, University of Southern Denmark, Sonderborg*
[2] *Mobile Dialog Systems, IIKT, Otto von Guericke University, Germany*
olni@sdu.dk, siegert@ovgu.de

Abstract: Emotions are an integral part of a speaker’s charismatic impact. Previous studies took started from this impact examined the associated emotional features on the part of the speaker and the recipient. We start here from the emotions themselves and test with a view to, e.g., everyday business communication and based on isolated, enacted stimulus sentences, which emotions make speakers sound more or less charismatic and how this interacts with speaker gender and speech compression. The results of a perception experiment with 21 listeners show that high-arousal emotions make speakers sound more charismatic than low-arousal emotions. Moreover, some compression codes, including the popular MP3 codec, perform surprisingly poorly at differentiating emotions in terms of perceived speaker charisma, particularly in combination with female speakers’ voices.

1 Introduction

Charismatic expressions are a complex compound of communicative and non-communicative stimulus signals. Antonakis et al. defines charisma as "values-based, symbolic, and emotion-laden leader signaling" [1, p. 303]. This definition is developed further in Michalsky et al.: charisma is a phenomenon based on three signal pillars: competence, self-confidence, and passion. Competence creates trust on the part of the recipient [2]. Self-confidence triggers motivation, and passion leads to inspiration and commitment. Furthermore, this definition of charisma implicitly places a stronger focus on the non-verbal communication signals of prosody and body language (also referred to as "delivery", [3]), which indeed have often proven to be more powerful than the word in the perception and modeling of speaker charisma [3, 4, 5]. How powerful the modulation of prosodic information regarding charismatic impact could be, even when the prosodic signals came from machines or robots rather than from human beings, has been demonstrated e.g. in [6] and [7]. These and further experiments proved that for the sensation cocktail called charisma all four dimensions of prosody are involved in charisma perception - (1) pitch, (2) duration/timing, (3) loudness, (4) voice quality - and each with a variety of their corresponding acoustic parameters [8, 9, 10]. Pitch, for example, is relevant in the form of the average pitch level, as well as additionally in the form of the pitch range, the pitch variability and the pitch minimum (at the end of conclusive statements), see [11].

Psychologists’ research on emotions has sought to determine the nature of emotions for a long time, starting from the description of emotions either in a categorial [12] or dimensional way [13]. From the finding of universal emotions [14, 15], up to formulating emotional components [16]. Furthermore, in psychological research it is common sense that emotions reflect short-term states, usually bound to a specific event, action, or object [17]. Hence, an emotion reflects a distinct user assessment related to a specific experience [18]. The "inner" state of an emotional reaction is then manifested in certain bodily response patterns, describing measurable changes in the nervous systems and derived changes in face, voice, and body, which then can be observed by either human beings or sensors of a technical system. Scientific studies aiming to identify vocal response patterns have examined them most commonly by decomposing the acoustic waveform of speech and afterwards, assessing whether such acoustic properties are associated with the emotional state of a speaker. Studies typically investigated acted basic emotions or real, but bipolar inductions (e.g. low vs. high stress) and reported correlations between arousal and pitch: higher levels of arousal have been linked to higher-pitched vocal samples [[19]. A broader study measured differences in acoustic changes of 14 emotions, including various intensities of the same emotion family (e.g., cold anger, hot anger) and 29 acoustic variables [18]. The authors found that a combination of ten acoustic properties distinguish discrete emotions to a greater extent than an attribution to valence and arousal alone. However, these links were complex and multivariate in nature, involving post-hoc comparisons. Furthermore, the acoustic characteristics are described mostly in a qualitative manner, such as "medium low frequency energy" or "increase of pitch over time".

Regarding the comparison of emotional speech and charismatic speech both as well as their correlations have been studied a lot [20, 21], but to the best of our knowledge there is no analysis present using the whole range of full-blown (discrete) emotions as research subject, especially not under psychophonetic aspects.

Another development that influences modern communication is the globalization of markets and societies. It has for years led to an increased use of digital communication tools like smartphones and, in particular, video-call systems like Zoom, Teams, Skype, WhatsApp, Facetime etc. This development has been dramatically accelerated by the COVID-19 pandemic. The number of video calls alone has increased by about 900 % during the past 15 years; 2020 even saw an additional temporary peak growth of 2,900% due to COVID-19 shutdowns [22, 23]. Companies like Cisco assume that in 2022 about 80% of the global internet traffic is caused by video calls [23]. From a charismatic and affective perspective speakers are primarily left with prosody as their only means of creating a charismatic effect or transmitting an affective reaction, as in video calls, body language and eye contact are often limited, the webcam has a disadvantageous viewing angle, and the transmitted video blurred or dark. Not to forget normal telephone calls and voice messages without a video signal.

On top of that charismatic and emotional expressions of a speaker are further influenced by the fact that all of these tools use audio compression in transmitting the speakers'voices. This is a potential problem per se – which becomes even bigger when the tools are used on the go, where the often reduced network quality leads stronger audio compression. Many studies have already addressed the question of how this audio compression affects the acoustic speech signal and its reception. However, the vast majority of these studies focus on questions of speech intelligibility or speaker identification [24, 25]. The speakers' perceived traits or impact on the audience have rarely been dealt with so far [26, 27, 28]. Siegert & Niebuhr [29, 30] have shown in two earlier studies that audio compression negatively affects perceived charisma as well as its acoustic foundation in a codec-specific manner, especially if the speaker is female. However, these earlier studies were carried out exclusively on the basis of emotionally neutral statements. This reduces the ecological validity of the findings insofar as everyday communication, and business communication in particular, is typically more often emotionally colored than neutral. We have therefore carried out here a further perception experiment that uses emotional instead of neutral stimuli and that tests the following three questions (Q1-3):

(Q1) At the interface of emotion and charisma, are the basic emotions anger, disgust, fear, joy, and sadness and boredom overall distinguishable along ten established charisma scales?

(Q2) How does a compression at low bitrates of four popular audio codecs influence this distinction in Q1?

(Q3) How does speaker gender interact with Q1 and Q2?

2 Methods

2.1 Utilized Speech Stimuli

This study made use of the same speech stimuli as in [29, 30] to maintain comparability: the Berlin Database of Emotional Speech (EMO-DB) [31] containing recordings of German sentences by 10 professional actors (five female). The sentences contain emotionally neutral verbal content but were uttered with different emotional expressions and in a neutral matter-of-fact version. EMO-DB comprises high-quality recordings in both technical and acoustic terms, stored as uncompressed WAV files (mono, sampling rate 16 kHz, 16-bit quantization depth, bit-rate 256 kBit/s). The high acoustic quality achieved by studio recordings of trained speakers with clear sonorous voices is one reason why this database is seen as a benchmark dataset for various applications [32]. For our study, we selected the same four speakers (male: #11 and #15 female: #13 and #14) as in the previous studies [29, 30], but with all emotional expressions including the neutral realization.

2.2 Utilized Compression Codecs

Speech compression for mobile communication is mostly used to reduce the bandwidth for transmission, the transmission delay as well as required system memory and storage [25, 33]. A numerous variety of different compression techniques for different applications is available, so that a selection of certain codecs necessary for an initial investigation is needed. To be in line with the previous investigations, the same four codecs were utilized.

Adaptive Multi-Rate Wideband (AMRWB) is a high-quality speech audio codec developed for mobile communication [34], also known as “HD Voice” and Voice over LTE (VoLTE) due to the processing of a wider speech bandwidth (50-6400/7000 Hz). The codec is based on Algebraic Code-Excited Linear Prediction (ACELP) and Linear Predictive Coding (LPC) parameters. We chose a bit-rate of 12.65 kBit/s, which is intended for pure speech signals [34].

MPEG-1/MPEG-2 Audio Layer III (MP3) is a well-known lossy compression codec specifically developed for music. By identifying and discarding those parts of the original sound signal that are assumed to exceed a listener’s auditory resolution ability, a perceptual coding is implemented. Besides its famous usage for music streaming, lower bitrates (16 kBit/s) are also used to encode audio dramas [35]. Moreover, MP3 has become so popular now also for data in the speech sciences that it is one of the file formats that PRAAT can process [36].

OPUS is an open-source lossy audio codec, offering a speech-oriented operation (SILK, similar to Speex) as well as a low-latency music compression mode (CELT, similar to MP3) [37]. Furthermore, OPUS offers a hybrid mode to improve the speech intelligibility at low bit rates by enriching the synthesized signal with characteristics represented by a psychoacoustic model. This hybrid mode is activated by specific bitrates, which is 34 kBit/s in the present study.

SPEEX is an open-source lossy speech codec [38]. It uses Code-Excited Linear Predictiop. Altough considered obsolete, it is still used as a speech transmission codec in some speech assistants [39]. We chose the lowest SPEEX quality parameter (quality=0, i.e. 3.95 kBit/s).

2.3 Listener Evaluation

The listener evaluation was conducted via an online survey tool (SoSci Survey Version 3.2.03-i, [40]). To avoid that similar samples (same speaker and/or encoding quality) directly follow each other, the samples were presented in a pseudo-randomized order. To ensure that the surveys can be technologically completed by all listeners and that the annotation task has been understood, a short audio test and an introduction of the variables was presented in the beginning. Afterwards, the listeners (aka labellers) were asked to listen to all stimuli subsequently and rate the respective speaker's performance on five-point Likert scales (ranging from 1 "not at all" to 5 "very strong"). Previous studies showed that speaker charisma is a fairly complex concept and not always easy to apply by listeners in rating tasks. Thus, in addition to a single assessment of the dimension charisma, we furthermore included nine other speaker attributes closely related to speaker charisma and derived from the studies of [8], [41] as well as from the charisma model of [42] and have already been successfully applied in previous studies of charisma rating tasks (e.g., [43, 44]). At the end of the rating task, socio-demographic information (age, sex, mother tongue, BFI-S16) was collected from the participants. Including this additional information, the entire perception experiment took about one hour.

2.4 Statistical analysis

The statistical analysis of the listeners' ratings along the seven emotion scales was carried out using a repeated-measures analysis of covariance (ANCOVA). The ANCOVA was based on the four fixed within-subject factors addressed in our research questions (Q1)-(Q3):

- *Emotion*: seven factor levels corresponding to basic emotions adopted from previous studies, i.e. fear, disgust, boredom, neutral, anger, joy, and sadness.
- *Scale*: ten factor levels, each representing a rating dimension that was successfully used in previous studies for the differentiated and sensitive evaluation perceived speaker charisma, i.e. visionary, inspiring, confident, charming, attractive, persuasive, charismatic, credible, passionate, decided.
- *Codec*: four factor levels corresponding to the popular codecs that there, at high compression rate, applied to the original sentence recordings, i.e. AMRWB, MP3, OPUS, and SPEEX.
- *Sex*: two factor levels associated with the two male and female speakers whose recorded sentences were the basis for the charisma ratings and the codec compression.

The two male and female speakers themselves constituted the covariate Speaker, which was included in the statistical model as a control. The minimum sample size at each factor level corresponds to the number of listeners, i.e. N = 21. Greenhouse-Geisser corrections were used in case the sphericity criterion was not met by the data. Effect sizes were estimated by means of partial eta-squared (η_p^2).

3 Results

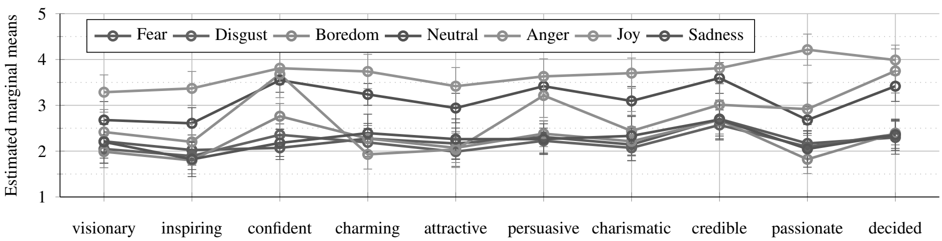

Figure 1 – Results summary in terms of the Scale*Emotion interaction. Each estimated marginal mean represents 336 listener ratings on a 1-5 Likert scale. Vertical error bars show 95% CIs.

ANCOVA results showed significant main effects for all fixed within-subject factors except Sex. In terms of effect sizes, the strongest main effect was that of Emotion ($F[6,120] = 29.193$, $p < 0.001$, $\eta_p^2 = 0.593$), closely followed by that of Codec ($F[3,60] = 28.241$, $p < 0.001$, $\eta_p^2 = 0.585$). The main effect of Scale was overall weaker ($F[1,20] = 9.166$, $p = 0.007$, $\eta_p^2 = 0.314$). The covariate Speaker turned out to be relevant neither alone nor in interaction with any of the four fixed within-subject factors. Only the three-way interaction Emotion*Codec*Speaker approached statistical significance. In that respect, Speaker differs from Sex. Although the latter also yielded no separate main effect, it interacted significantly in a number of ways with the other fixed within-subject factors. The two most noticeable interactions in terms of effect sizes were Sex*Emotion (F[6,120] = 6.333, $p < 0.001$, $\eta_p^2 = 0.241$) and Sex*Codec (F[3,60] = 18.270, $p < 0.001$, $\eta_p^2 = 0.447$) as well as the three-way interaction Sex*Emotion*Codec (F[18,360] = 10.773, $p < 0.001$, $\eta_p^2 = 0.350$).

Regarding the remaining pattern of interactions, note that Scale*Emotion was far from being significant, while the three-way interaction with Codec was significant (Scale*Emotion*Codec: F[18,360] = 5.375, $p < 0.001$, $\eta_p^2 = 0.212$). The two-way interaction Emotion*Codec was significant too (F[18,360] = 15.499, $p < 0.001$, $\eta_p^2 = 0.437$) and stronger in terms of effect size than Scale*Emotion*Codec. All other interactions came out not significant or only approached significance. The latter also concerned the interaction among the four fixed within-subject factors. Due to the brevity of this research report, we will focus on those points of the results that are most relevant to questions (Q1)-(Q3). Further two-way statistical analyses were conducted to that end, and pairwise comparisons (with Sidak corrections of alpha-error levels) were carried out to search for significant difference among factor levels. All these additional test statistics cannot be reported here in detail, but they provided the basis for the following in-depth descriptions and illustrations of key results.

Figure 1 summarizes the results on the extent to which the established ten charisma scales were relevant and sensitive beyond measuring charisma; in particular, whether they were able to significantly differentiate between basic emotional states of the speaker (Q1). Indeed, as we see in Figure 1, the tested basic emotions led to different rating levels across the ten scales. Joy achieved the highest ratings on each scale. That is, when the speakers realized the individual sentences with joy, they sounded significantly more visionary, inspiring, confident etc. than with all other emotions. After joy followed, with overall significantly lower rating levels, already the non-emotional speaker state 'neutral', which, in turn, was followed by anger, again with a significant step down in rating levels on many scales – but not on all scales. That is, in statistical terms, angry speakers sounded equally well in the ears of listeners on a number of charisma-related scales as the emotionally neutral speakers. This concerned the scales confident, persuasive, passionate, and decided. On the charisma scale itself, however, angry speakers were rated significantly lower than emotionally neutral speakers. The significantly lowest overall ratings were obtained by the sentences realized with disgust. The three emotions in between anger and disgust – i.e. fear, boredom, and sadness – were statistically indistinguishable as to their ratings along the ten charisma scales. So, it was neither the bored nor sad speaker who sounded least charismatic, it was the disgusted speaker. We also see in Figure 1 that the ups and downs in rating levels across the ten scales were for all emotions largely similar, which is why we found no significant interaction of Scale*Emotion.

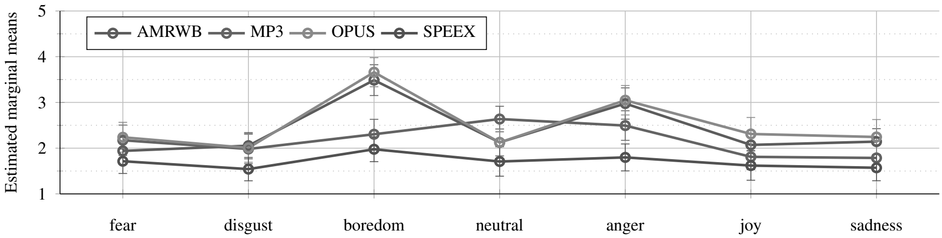

Figure 2 – Results summary in terms of the Emotion*Codec interaction. Each estimated marginal mean represents 840 listener ratings. Vertical error bars show 95% CIs

With regard to (Q2) and the effects of Codec, key results are summarized in Figure 2. The most obvious result was that, with the SPEEX codec, all emotions obtained the significantly lowest ratings across all ten charisma scales; and, perhaps even more importantly, all emotions were rated statistically equally low. That is, emotional speaker states, including strongly different ones like joy, fear, anger, and disgust, became indistinguishable for listeners under SPEEX, at least in terms of the impact and performance ratings covered by the ten charisma scales. In other words, the speakers were always perceived as equally (un)charismatic, no matter whether they spoke, for example, joyfully, angrily, disgusted, or with strong fear in their voice. This was true for no other codec than SPEEX (cf. the Emotion*Codec interaction above). The statistically most differentiated emotion perception along the charisma scales was possible using AMRWB and OPUS. Interestingly, the popular MP3 standard performed significantly worse. Unlike AMRWB and OPUS, MP3 made listeners not discriminate boredom from fear and anger from the emotionally neutral speaker state. Furthermore, in terms of how much the codecs interfered with the speaker's charismatic impact, we found that OPUS made speakers sound overall most charismatic (i.e. across all emotions), followed by AMRWB, MP3, and SPEEX, each of which caused a significant step down in perceived speaker charisma relative to the preceding one.

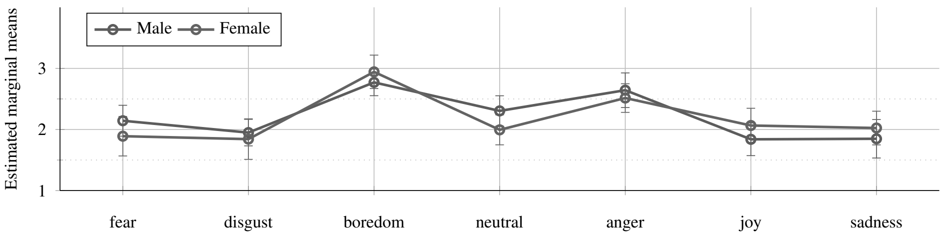

Figure 3 – Results summary in terms of the Emotion*Sex interaction. Each estimated marginal mean represents 1,680 listener ratings. Vertical error bars show 95% CIs.

The effects of Sex addressed by (Q3) manifested themselves at many points in the results. Sex is, before Codec, the factor involved in most of the significant interactions, including Sex*Emotion and Sex*Codec. Thus, emotion perception and codec compression were both very sensitive to the difference between male and female speakers' voices. Figures 3 and 1 display aspects of the corresponding key results. Figure 3 relates to the Sex*Emotion interaction. We see that Sex did not determine which emotions were well distinguishable and which were not on the given ten scales. However, we see that the female speakers received overall lower ratings, i.e. sounded significantly less charismatic than the male speakers, when speaking with fear, disgust, or anger. Even in the emotionally neutral state, women were rated to perform significantly worse than men. The opposite was true for boredom and, in particular, for joy and sadness. In combination with these three emotions, women's voices yielded a significantly higher (charisma) rating level than the men's voices.

Regarding the interaction of Sex with the Codec, Figure 1 shows that, for the tested high compression rates, OPUS was not only the one that made speakers sound the most charismatic overall along the ten scales. It was also the only codec that treated males' and females' voices equally well averaged across all emotions in this regard. Under MP3, the women's voices lost significantly and disproportionally more of their overall charismatic impact than the men's voices. The opposite was true for AMRWB and – although the difference was relatively small – also for SPEEX. For these two codes, the women came off overall significantly better, i.e. retained more of the charismatic impact in their voices than men. Breaking down these sex-specific codec performances according to the individual emotions (cf. the above Sex*Emotion*Codec interaction) additionally revealed that the disadvantage of MP3 for women was primarily due to the emotionally neutral speaker state as well as to the two diametrically opposed basic emotions of anger and joy. Interestingly, the advantage women had in terms of their ratings under AMRWB was also due to the two emotions anger and joy. Moreover, women obtained higher ratings in combination with disgust, too. The advantage women had in connection with SPEEX was linked to anger as well as to fear. Finally, what is not shown in Figure 1 is that OPUS gave men a significant (charisma) rating advantage in expressing two emotions: fear and boredom.

4 Discussion and Conclusions

Three questions were put forward at the outset of this paper. (Q1) asked whether and to what degree a set of ten Likert scales that reliably measure levels of perceived speaker charisma can also be used to differentiate basic emotional speaker states. Our data gave a partially positive answer to that question. However, they also suggested that the set of 10 scales was better at discriminating emotions along their arousal dimension than along their valence dimension. Accordingly, speakers (of both sexes) with high-arousal emotions such as anger and joy were among the top performers on the charisma scales (despite the diametrically opposed valences of anger and joy), whereas speakers with low-arousal emotions such as sadness, fear, and boredom all sounded similarly less charismatic. The particular sensitivity of the scale set to arousal is plausible given that valence is largely irrelevant to the measurement of a concept like charisma whose definition relies on positive attributes. Sad and angry speakers are also typically not included (explicitly) in speaker samples of charisma studies. Nevertheless, our present findings raise the question for future studies whether angry speakers can indeed come across as charismatic as joyful speakers (some political leaders make us think so) and whether sad or anxious speakers can be as charismatic as speakers who are just bored, cf. also [45, 46]. Alternatively, established set of charisma scales could also be too blind along the valence dimension to adequately reflect existing emotion-charisma interactions on the part of the audience.

The results on this future research question will also affect the interpretation of the present results on (Q2), which concerned the interaction of emotions and speech compression codecs. Assuming that the set of 10 scales is necessary and sufficient to measure an audience's perception of charisma, the key result of our study is as follows: Speech compression affects the acoustic cues to emotions in such a way that the assessment of speaker charisma becomes blind to certain emotional states of the speaker. In other words, under certain compression conditions, it is not disadvantageous for the speaker to be angry rather than neutral, or to be anxious instead of just bored. Emotions like anger and fear are sort of "filtered out" of the signal. It is amazing that SPEEX neutralized practically all emotional differences expressed by speakers (of both sexes). Thereby, it actually completely disqualifies itself for use in everyday digital communication. Also amazing is the poor performance of the MP3 codec. With regard to the strongly reduced level of perceived speaker charisma found for MP3, our findings are a replication of Siegert & Niebuhr [29]. What is new here is the finding that MP3 also turned out to be a clearly underperforming codec with regard to the differentiation of emotional speaker states. The female voices in particular suffered from MP3, both in terms of emotion and of charisma perception. In contrast, regarding these perceptions, OPUS was not only the best tested codec, but also a largely sex-neutral one. So if men and women want to express the full spectrum of their emotions in digital communication, then OPUS should be the speech compression codec of choice. However, if women do not want that, for example, because they are unsure whether fear (e.g., due to public-speaking anxiety) or boredom occur frequently in their digital communication and reduce their charismatic impact, then AMRWB is the better codec for them. Men should continue to use OPUS for the same reasons. These practical conclusions are also the main answers to (Q3).

To conclude, our results show the importance of comparative research into the performance of different speech compression codecs in relation to the non-verbal forms and functions of everyday communication. Our results also show what practical implications can arise from such research. In addition to the further research questions discussed in connection with (Q1) above, we will next investigate in more detail at which compression rate for which codec the negative effects on emotions and charisma begin to unfold, again taking a sex-sensitive perspective.

Acknowledgements

This research is kindly supported by the IE-Industrial Elektronik project (SFD-17-0036) which has received EU co-financing from the European Social Fund. We would furthermore like to thank Krestina Vendelbo Christensen and Radek Skarnitzl for inspiring discussion about future research on the socio-phonetic issues of compressed speech.

References

[1] ANTONAKIS, J., N. BASTARDOZ, P. JACQUART, and B. SHAMIR: *Charisma: An ill-defined and ill-measured gift. Annual Review of Organizational Psychology and Organizational Behavior*, 3, pp. 293–319, 2016.

[2] MICHALSKY, J. and O. NIEBUHR: *Myth busted?: challenging what we think we know about charismatic speech. Acta Universitatis Carolinae*, (2), pp. 27–56, 2019. doi:10.14712/24646830.2019.17.

[3] CASPI, A., R. BOGLER, and O. TZUMAN: *"judging a book by its cover": The dominance of delivery over content when perceiving charisma. Group & Organization Management*, 44(6), pp. 1067–1098, 2019. doi:10.1177/1059601119835982.

[4] SCHERER, S., G. LAYHER, J. KANE, H. NEUMANN, and N. CAMPBELL: *An audiovisual political speech analysis incorporating eye-tracking and perception data*. In *Proc. of the LREC'12*, pp. 1114–1120. ELRA, Istanbul, Turkey, 2012.

[5] WÖRTWEIN, T., M. CHOLLET, B. SCHAUERTE, L.-P. MORENCY, R. STIEFELHAGEN, and S. SCHERER: *Multimodal public speaking performance assessment*. In *Proc. of the ACM ICMI'15*, p. 43–50. New York, NY, USA, 2015. doi:10.1145/2818346.2820762.

[6] FISCHER, K., O. NIEBUHR, L. C. JENSEN, and L. BODENHAGEN: *Speech melody matters—how robots profit from using charismatic speech. J. Hum.-Robot Interact.*, 9(1), 2019.

[7] NIEBUHR, O. and J. MICHALSKY: *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 *Computational Science and Its Applications – ICCSA 2019*, vol. 11620 LNCS, pp. 375–390. Springer, Saint Petersburg, 2019. doi:10.1007/978-3-030-24296-1_31.

[8] ROSENBERG, A. and J. HIRSCHBERG: *Charisma perception from text and speech. Speech Communication*, 51(7), pp. 640–655, 2009.

[9] SIGNORELLO, R., F. D'ERRICO, I. POGGI, and D. DEMOLIN: *How charisma is perceived from speech: A multidimensional approach*. In *Proc. of the ASE/IEEE SOCIALCOM-PASSAT '12*, pp. 435–440. 2012. doi:10.1109/SocialCom-PASSAT.2012.68.

[10] NIEBUHR, O., R. SKARNITZL, and L. TYLEČKOVÁ: *The acoustic fingerprint of a charismatic voice - initial evidence from correlations between long-term spectral features and listener ratings*. In *Speech Prosody'18*, pp. 359–363. 2018.

[11] NIEBUHR, O. and R. SKARNITZL: *Measuring a speaker's acoustic correlates of pitch - but which? a contrastive analysis for perceived speaker charisma*. In *Proc. of the 19th International Congress of Phonetic Sciences,*, pp. 1774–1778. Melbourne, Australia, 2019.

[12] McDOUGALL, W.: *An introduction to Social Psychology*. Methuen & Co, London, UK, 2 edn., 1908.

[13] WUNDT, W. M.: *Vorlesungen über die Menschen- und Tierseele*. L. Voss, Leipzig, 6 edn., 1919.

[14] PLUTCHIK, R.: *Emotion, a psychoevolutionary synthesis*. Harper & Row, New York, USA, 1980.

[15] EKMAN, P.: *Basic Emotions*, pp. 45–60. John Wiley & Sons, Hoboken, USA, 2005.

[16] SCHLOSBERG, H.: *Three dimensions of emotion*. *Psychol Rev*, 61, pp. 81–88, 1954.

[17] BECKER, P.: *Structural and relational analyses of emotions and personality traits*. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 22, pp. 155–172, 2001.

[18] BANSE, R. and K. R. SCHERER: *Acoustic profiles in vocal emotion expression*. *J Pers Soc Psychol*, 70, pp. 614–636, 1996.

[19] BACHOROWSKI, J. A.: *Vocal expression and perception of emotion*. *Curr Dir Psychol Sci*, 8, pp. 53–57, 1999.

[20] BONO, J. E. and R. ILIES: *Charisma, positive emotions and mood contagion*. *The Leadership Quarterly*, 17(4), pp. 317–334, 2006. doi:10.1016/j.leaqua.2006.04.008.

[21] WASIELEWSKI, P. L.: *The emotional basis of charisma**. *Symbolic Interaction*, 8(2), pp. 207–222, 1985. doi:10.1525/si.1985.8.2.207.

[22] SCOTT, R.: *Must have video conferencing statistics 2020*. In *UCToday*. 2020. URL `http://bit.ly/UCToday-VideoConferencing2020`. [Online; posted 27-Jul-2020].

[23] HARGRAVE, S.: *In the grip of a climate crisis, demand for video calls is soaring*. In *Wired*. 2020. URL `http://bit.ly/Wired-VideoCalls2020`. [Online; posted 27-Jan-2020].

[24] BYRNE, C. and P. FOULKES: *The 'mobile phone effect' on vowel formants*. *International Journal of Speech, Language and the Law*, 11(1), pp. 83–102, 2004.

[25] MARUSCHKE, M., O. JOKISCH, M. MESZAROS, F. TROJAHN, and M. HOFFMANN: *Quality assessment of two fullband audio codecs supporting real-time communication*. In *Proc. of the 18th International Conference on Speech and Computer SPECOM 2016*, pp. 571–579. 2016.

[26] DOBBRINER, J., O. JOKISCH, and M. MARUSCHKE: *Assessment of prosodic attributes in codec-compressed speech*. In *Proc. 12th Conference Phonetik und Phonologie im deutschsprachigen Raum (P&P)*, pp. 35–39. 2016.

[27] SIEGERT, I., A. F. LOTZ, M. MARUSCHKE, O. JOKISCH, and A. WENDEMUTH: *Emotion intelligibility within codec-compressed and reduced bandwith speech*. In *ITG-Fb. 267: Speech Communication : 12. ITG-Fachtagung Sprachkommunikation*, pp. 215–219. 2016.

[28] SIEGERT, I., A. F. LOTZ, O. EGOROW, and A. WENDEMUTH: *Improving Speech-Based Emotion Recognition by Using Psychoacoustic Modeling and Analysis-by-Synthesis*, pp. 445–455. Springer International Publishing, Cham, 2017. doi:10.1007/978-3-319-66429-3_44.

[29] SIEGERT, I. and O. NIEBUHR: *Women, be aware that your vocal charisma can dwindle in remote meetings*. *Frontiers in Communication*, 5, p. 7, 2021. doi:10.3389/fcomm.2019.00012.

[30] SIEGERT, I. and O. NIEBUHR: *Speech signal compression deteriorates acoustic cues to perceived speaker charisma*. In *Elektronische Sprachsignalverarbeitung 2021. Tagungsband der 32. Konferenz*, vol. 99 of *Studientexte zur Sprachkommunikation*, pp. 1–10. TUDpress, Berlin, Germany, 2021.

[31] BURKHARDT, F., A. PAESCHKE, M. ROLFES, W. SENDLMEIER, and B. WEISS: *A database of german emotional speech*. In *Proc. of the INTERSPEECH*, pp. 1517–1520. 2005.

[32] SCHULLER, B., B. VLASENKO, F. EYBEN, G. RIGOLL, and A. WENDEMUTH: *Acoustic Emotion Recognition: A Benchmark Comparison of Performances*. In *Proc. of the IEEE ASRU-2009*, pp. 552–557. 2009.

[33] SIEGERT, I., A. F. LOTZ, L. L. DUONG, and A. WENDEMUTH: *Measuring the impact of audio compression on the spectral quality of speech data*. In *Elektronische Sprachsignalverarbeitung 2016*, vol. 81 of *Studientexte zur Sprachkommunikation*, pp. 229–236. Leipzig, Germany, 2016.

[34] ITU-T: *Wideband Coding of Speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)*. REC G.722.2, 2003. URL `https://www.itu.int/rec/T-REC-G.722.2-200307-I/en`.

[35] AHERN, S.: *Acoustical design of concert halls and theatres: a personal account*. Routledge, 2020.

[36] BOERSMA, P.: *Praat, a system for doing phonetics by computer*. *Glot Int.*, 5, pp. 341–345, 2001.

[37] VALIN, J.-M., K. VOS, and T. TERRIBERRY: *Definition of the opus audio codec*. RFC 6716, 2012. URL `http://tools.ietf.org/html/rfc6716`.

[38] XIPH.ORG FOUNDATION: *Speex: A free codec for free speech*. 2014. URL `http://www.speex.org/software/`.

[39] CAVIGLIONE, L.: *A first look at traffic patterns of Siri*. *Transactions on Emerging Telecommunications Technologies*, 26(4), pp. 664–669, 2015. doi:10.1002/ett.2697.

[40] LEINER, D. J.: *Sosci survey (version 3.1.06)*. 2019. Available at https://www.soscisurvey.de.

[41] WENINGER, F., B. SCHULLER, A. BATLINER, S. STEIDL, and D. SEPPI: *Recognition of nonprototypical emotions in reverberated and noisy speech by nonnegative matrix factorization*. *EURASIP Journal on Advances in Signal Processing*, 2011(1), p. 838790, 2011.

[42] MICHALSKY, J.: *Einfach überzeugend! - Teil 1 - Mehr Charisma durch das Verständnis von Sprache und Stimme*. Virtuelle Hochschile Bayern (VHB), 2019. URL `https://open.vhb.org/blocks/ildmetaselect/detailpage.php?id=90&lang=en`.

[43] NIEBUHR, O. and S. WRZECZSZ: *A woman's gotta do what a woman's gotta do, and a man's gotta say what a man's gotta say - sex-specific differences in the production and perception of persuasive power*. In *Proc. 16th International Pragmatics Conference, Hong Kong, China*, pp. 1–2. 2019.

[44] NIEBUHR, O.: *Space fighters on stage - how the f1 and f2 vowel-space dimensions contribute to perceived speaker charisma*. In R. BÖCK, I. SIEGERT, and A. WENDEMUTH (eds.), *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2020*, pp. 265–277. TUDpress, Dresden, 2020.

[45] D'ERRICO, F.: *'too humble and sad': The effect of humility and emotional display when a politician talks about a moral issue*. *Social Science Information*, 58(4), pp. 660–680, 2019. doi:10.1177/0539018419893564.

[46] D'ERRICO, F., O. NIEBUHR, and I. POGGI: *Humble voices in political communication: A speech analysis across two cultures*. In *Computational Science and Its Applications – ICCSA 2019*, pp. 361–374. Springer, Cham, 2019.