

# Winning Over an Audience – A Perception-based Analysis of Prosodic Features of Charismatic Speech

Stephanie Berger<sup>1</sup>, Oliver Niebuhr<sup>2</sup>, Benno Peters<sup>1</sup>

<sup>1</sup> *ISFAS – Abteilung für Allgemeine Sprachwissenschaft, CAU Kiel, E-Mail: stephanie.berger@freenet.de*

<sup>2</sup> *Mads Clausen Institute, Technology Entrepreneurship and Innovation, University of Southern Denmark*

## Introduction

Charisma is a term that is difficult to define, and yet it is a widely used concept in today's western societies since it plays a role in public presentations of business strategies, new products, ideas for the future of a political party, country, or city, or for personalities in the entertainment industry. While many people are successful with their ideas, products or programs alone, some are captivating public speakers that create a following not solely by their visions, but also by their personality. These people are known as charismatic leaders or charismatic personalities.

The definition of charisma has evolved over time. Max Weber was the first to use the term in a non-religious setting and defined charisma as "a certain quality of an individual personality by virtue of which [the speaker] is considered extra-ordinary and treated as endowed with supernatural, superhuman, or at least specifically exceptional powers or qualities" [1]. Nowadays, charisma is mostly seen as a concept that stems from the relationship between a leader who has qualities that are perceived as charismatic, followers who have to be willing to follow a leader and should already share the leader's mindset and vision, and an environment that fosters charisma which does not necessarily need to be a crisis, but can simply be a situation of uncertainty (see [2], but also [3]).

The charismatic effect of a leader is created by many different elements that can be classified into two areas: the content of the message (the language of the message, its structure as well as the vision that is presented) and its delivery (where voice, posture, gestures, body language etc. belong to). Speakers with a strong delivery are perceived to be more charismatic [4]. These aspects result in a charisma concept that is gradual rather than 'all or nothing'. It is created by a combination of different aspects like leader qualities, follower ideas and specific contexts, while being influential is not a necessity [5], but "an individual can be charismatic without having any influence whatsoever" [3, p. 304].

The present study only deals with a few elements of the delivery-related qualities of a charismatic speaker: The acoustic parameters fundamental frequency (F0) level, F0 range and speech rate and their effect on the rating of a speaker as being charismatic will be investigated by means of a perception experiment. Of course, there are many more elements to charisma, including aspects that are not related to the voice, which will not be addressed further here. This paper will first briefly review the previous phonetic research on charismatic speech, before the methods of the present study are introduced. The results are then presented, before they are discussed.

## Previous research

The majority of phonetic research on charismatic speech that has been published so far has focused on political speech (see, for example, [6, 7, 8]), but studies on other areas like business-related speeches begin to appear as well [5, 9].

Rosenberg and Hirschberg created perception experiments where native speakers of American English rated speech excerpts from political speeches on five-point Likert scales with statements like 'The speaker is charismatic/charming/convincing/etc.' [6, 7]. In terms of F0 level, F0 range and speech rate, the results suggested that a stimulus with higher F0 level was perceived as more charismatic. The same was the case for a faster speech rate and a larger F0 range [6, 7].

The study of [8] suggests similar preferred acoustic features for charismatic speech. Their study focuses on cultural differences of the perception of charismatic speech. Their results for American listeners rating American English speakers suggest the same as [6, 7] above: Speakers with a faster speech rate, higher F0 level and larger pitch variation were rated as more charismatic. For a more detailed discussion and an overview of the cultural differences of charisma perception, the reader is referred to the original paper [8].

In the business context, [5] and [9] analyzed a large number of acoustic features from product presentations of Apple's co-founder and former CEO Steve Jobs – who is often considered as a very charismatic speaker – and Facebook's CEO Mark Zuckerberg – often perceived as a non-charismatic speaker. The results of the acoustical analyses were compared to reference values from the literature. Their results on F0 level, F0 range and speech rate suggest that Steve Jobs had a very high F0 level and a larger F0 range – compared to the reference sample at the lower end of female F0 levels and an around six semitones larger F0 range than the reference values. Steve Jobs was also found to have a fairly high, but not exceptionally high, speech rate. These features were also found for Mark Zuckerberg, who also stands out from the reference sample, but not as clearly as Steve Jobs. A difference between the two speakers is the speech rate, as Mark Zuckerberg's speech rate clearly exceeded the range of the reference values, which "causes many strong phonetic reductions" [9, p. 2]. These results suggest that speech is perceived as charismatic when the speaker has a rather high F0 level, a large pitch range and fast, but not too fast, speech rate. Of course, F0 level, F0 range and speech rate are only three of many acoustic features that play together in the creation of a charismatic voice, and further aspects were also investigated in studies [5] and [9] in more detail.

The present study differs from the previous studies insofar that it tries to investigate tendencies of what characteristic the three parameters F0 level, F0 range and speech rate should exhibit (should they be on a higher, lower or intermediate level) in order to create the impression of a charismatic voice. This is done by using stimuli in a listening experiment that are digitally altered versions of the same sentence for two speakers which eliminates some elements that could influence listeners' ratings. The content of the sentence remained constant across both speakers and all manipulations which means that listeners could not side with the message of one speaker more than the other. On top of that, the general prosodic contour remained constant for each speaker as well which would not be possible with naturally spoken stimuli. The method will be addressed further in the sections below.

## Methods

### The speakers and the materials

For the present study, two male experienced native speakers of American English read the same text. Both speakers are lecturers at Kiel University and are therefore used to speak in front of classes. They come from roughly the same dialect area in the USA (Iowa and Minnesota). The recordings were made in the soundproof recording studio at Kiel University with a RME Fireface 800 audio interface and a Microtech Gefell M940 condenser microphone with super-cardioid directionality.

The speakers were asked to read an online article on the success of blockbuster films, in particular superhero films. The article was chosen because it was unusual enough to be interesting to read, but it still offered the possibility to choose a target phrase that could be interpreted as coming from the business, marketing or entertainment context: *'Let's see if that red logo above the title can draw people in.'* The participants of the experiment were not told that the sentence they listened to was related to business. The advantage of using the same phrase for both speakers is that the content of the message does not differ and therefore cannot influence charisma ratings as easily. The text was set in 14pt Times New Roman with 1.5 line spacing and justified print to resemble the research in [10] in order to elicit relatively expressive speech. For the same reason, the speakers were asked to pretend they would present the article to a class and were standing during the recordings.

### The stimuli manipulations

The stimuli for the perception experiment were manipulated in terms of F0 level, F0 range and speech rate using the PSOLA resynthesis algorithm [11] in Praat [12]. The original phrases were also stimuli in the experiment. They were, however, opened in Praat's manipulation editor and resynthesized to obtain a quality similar to the manipulated stimuli, as the resynthesis can create acoustic artifacts or change the sound quality of the recording.

The features F0 level, F0 range (FOR) and speech rate (SR) had three characteristics each: increased (HF0/HFOR/HSR), not altered (OF0/OFOR/OSR) and decreased (LF0/LFOR/

LSR). The stimuli were manipulated in all possible parameter combinations with some stimuli having only one altered acoustic feature, others two, yet others three. This finally resulted in 27 stimuli per speaker and 54 stimuli in total.

The pitch frequencies of the phrase were shifted either up or down by 2 semitones to change the F0 level. The F0 range was increased or decreased by using the formula in (1).

$$(1) \quad (\text{self} - \text{mean}) * \text{factor} - \text{mean}$$

The variable 'self' refers to the F0 value at each point of the pitch contour, from which the mean F0 value of the phrase is subtracted. Then, the difference is multiplied by a factor and subsequently, the mean was again subtracted from the solution to receive the value that would be added on to the value of self. For an increase of the F0 range, a factor of 1.5 was chosen, while a factor of 0.5 was used for the decrease of the range. The mean F0 value was adjusted to the characteristics of the F0 level. The speech rate was increased and decreased by one syllable per second. Table 1 below summarizes the factors used for the manipulations.

**Table 1:** The factors used for the stimuli manipulations.

Manipulation	F0 level	F0 range	Speech rate
higher	+ 2 st	factor: 1.5	+ 1 syll/s
lower	- 2 st	factor: 0.5	- 1 syll/s

### Experiment design

The perception experiment at the heart of this study was created and administered online using the website SoSci-Survey.de [13]. The experiment took about 30 to 40 minutes to complete. 16 participants (eight male, eight female; all native speakers of American English) completed the survey. The 54 stimuli were presented four times each with four different five-point Likert scales ranging from "strongly disagree" (= 5) through "neither agree nor disagree" (= 3) to "strongly agree" (= 1). The participants were asked to rate the speaker with each rendition, which resulted in 216 questions in the main part of the survey. Before and after this main part of the study, the participants were asked to provide relevant personal information on a voluntary basis.

The statements for the Likert scales can be found in examples (2) to (5). The abbreviated names of the scales are indicated in parentheses.

- (2) The speaker is charismatic. (CHA)
- (3) The speaker sounds attractive. (ATT)
- (4) The speaker is convincing. (CON)
- (5) The speaker is motivating. (MOT)

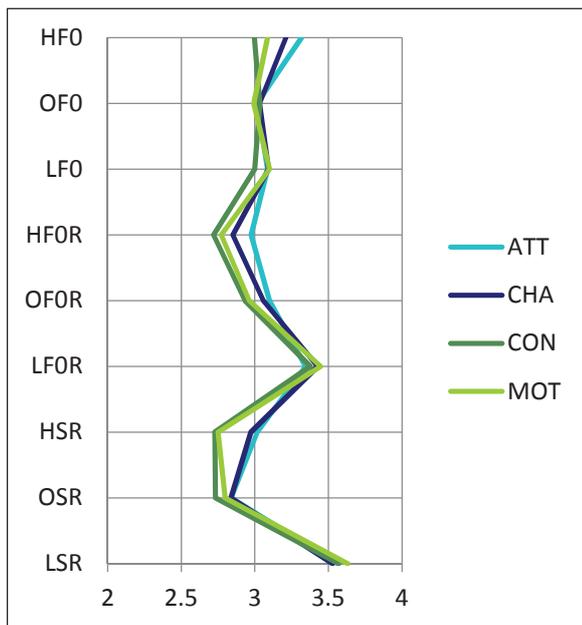
In addition to the term *charismatic*, the adjectives *attractive*, *convincing* and *motivating* were chosen because they are seen as closely connected to charisma in the literature (for example, [4, 6, 7]). These adjectives were thought to supplement the term *charisma*, as the concepts can be seen as perhaps supporting the charisma perception, but are not necessarily inherent characteristics of a charismatic person.

Each participant rated the scales and stimuli in the same order. The scales were presented nine times in a row with the stimuli randomized across each scale.

## Results

The ratings for each stimulus and scale were statistically analyzed using both explorative and inferential methods. For the visual exploration of the data, rating profiles as in Figure 1 were created. Furthermore, one-way ANOVAs for each parameter and scale and Tukey HSD post-hoc tests were calculated using RStudio [14]. For both statistical analyses, the probable influence of the combination of parameter characteristics on the rating of each stimulus were neglected. Each stimulus was represented once in every ANOVA depending on the characteristic of the parameter: once as either HF0, OF0 or LF0, once as either HF0R, OF0R or LF0R, and once as either HSR, OSR or LSR.

Figure 1 shows the rating profile of the mean ratings of the acoustic parameters on all four scales. It has to be kept in mind that the answers are grouped around the neutral rating possibility, which suggests that there were relatively few extreme ratings in the sample or that different groups within the sample levelled each other out. Table 2 provides the means of the Likert ratings and the corresponding standard deviations for each parameter.



**Figure 1:** Rating profile of the mean ratings of the stimuli. Each stimulus is represented three times – once for F0 level, F0 range (F0R) and speech rate (SR).

Unchanged and decreased F0 do not seem to have an influence on either of the scales. Figure 1 suggests that there was an effect regarding increased F0 level which is partially supported by the ANOVAs. Only the ANOVA for the attractiveness scale returned a highly significant main effect of F0 level ( $F[2,861] = 7.331$ ,  $p < 0.001$ ), there were no effects on the other scales. The post-hoc tests revealed that HF0 was rated to be less attractive than LF0 ( $p = 0.015$ ) and OF0 ( $p < 0.001$ ), but that there was no significant difference between LF0 and OF0.

**Table 2:** The mean values of the Likert ratings per scale for each acoustic parameter and their standard deviations (in parentheses).

	CHA	ATT	CON	MOT
HF0	3.21 (1.08)	3.32 (0.97)	3.00 (1.11)	3.09 (1.06)
OF0	3.03 (1.08)	3.01 (1.00)	3.04 (1.08)	2.99 (1.09)
LF0	3.09 (1.07)	3.09 (1.03)	3.00 (1.12)	3.10 (1.01)
HF0R	2.85 (1.09)	2.98 (1.01)	2.72 (1.11)	2.77 (1.03)
OF0R	3.06 (1.08)	3.10 (1.03)	2.93 (1.09)	2.96 (0.98)
LF0R	3.43 (0.98)	3.34 (0.95)	3.38 (1.01)	3.45 (1.05)
HSR	2.97 (1.02)	3.02 (0.91)	2.73 (1.07)	2.75 (0.98)
OSR	2.84 (1.01)	2.84 (0.96)	2.73 (0.97)	2.80 (0.96)
LSR	3.53 (1.07)	3.56 (1.01)	3.57 (1.05)	3.63 (1.00)

The visual inspection of Figure 1 suggests that an increased F0 range is judged more positively and decreased F0 range is judged rather negatively on all scales. This is supported by the ANOVAs that returned highly significant main effects on all four scales (ATT scale:  $F[2,861] = 10.148$ ; CHA scale:  $F[2,861] = 22.21$ ; CON scale:  $F[2,861] = 28.473$ ; MOT scale:  $F[2,861] = 33.853$ ;  $p < 0.001$  on all scales). While HF0R was rated significantly more positive than OF0R on the charismatic and convincing scales (CHA scale:  $p = 0.046$ ; CON scale:  $p = 0.042$ ), the differences were not significant on the attractive and motivating scales. The differences between the mean ratings of LF0R and OF0R were highly significant ( $p = 0.01$  on the ATT scale;  $p < 0.001$  on the CHA, CON and MOT scales), as were the differences between LF0R and HF0R ( $p < 0.001$  on all scales). These results combined with the results from Figure 1 suggest that a monotonous voice is not perceived as charismatic (and attractive, convincing and motivating). Pitch variety, however, seems to have a positive effect on perceived charisma, as was suggested by [5, 6, 7, 9].

Low speech rate was judged as most negative on all scales. Unchanged speech rate received on average quite positive judgements. The ANOVAs returned highly significant main effects for speech rate on all scales (ATT scale:  $F[2,861] = 44.225$ ; CHA scale:  $F[2,861] = 36.023$ ; CON scale:  $F[2,861] = 63.457$ ; MOT scale:  $F[2,861] = 74.203$ ;  $p < 0.001$  on all scales). There were no significant differences between HSR and OSR on either of the scales. The graphs for the ATT and CHA scales in Figure 1 might suggest a tendency towards a preference of a speech rate that is not too high, as was also suggested by [9]. Only the difference between LSR and HSR as well LSR and OSR was highly significant ( $p < 0.001$  for all scales and for both pairwise comparisons). These results indicate that the slow speech rate is perceived very negatively and a faster speech rate should be attempted in order to appear charismatic, which is also in line with previous research.

## Discussion and conclusion

While the acoustic features investigated for this study offer by no means an exhaustive idea of a charismatic voice, the present study allows a more controlled insight into the features that might play a role in creating the perception of a charismatic speaker. The adjectives used in this study seem to represent different concepts that perhaps support

charisma, but they cannot be seen as inherent aspects of charisma, since there were striking rating differences.

The results of the charismatic scale found on the basis of the perception experiment largely support the characteristics of the acoustic features of charismatic speech found in the literature. As was suggested by [5, 6, 7, 8, 9], a larger F0 range – meaning a greater pitch variety – and a faster speech rate seem to be preferred for charismatic speech, but there are also tendencies visible that suggest that the speech rate should not be too fast. However, in the present study, while the difference between unchanged and increased speech rate was not significant, there might have been a slight influence of the manipulations as the tempo was increased, but the degree of phonetic reduction was not adjusted.

In terms of F0 level, the results of this study cannot replicate the results of the previous studies. While previous research found a positive correlation between high F0 and charisma ratings, there were no significant effects of fundamental frequency in this study. The only significant effect of F0 was that higher F0 was perceived as less attractive, which is in line with results from attractiveness research (see [15]), but not charisma research. These results might suggest that F0 is not so much a feature that induces the perception of charisma, but rather either supports other features or does not play as much of a role in general. However, further research is needed to find an answer. There were rating similarities between convincing and charismatic, which might either suggest that these two concepts are related more closely than the others, or that they are perceptually differentiated in the basis of further acoustic parameters.

Working with synthesized stimuli for a charisma perception study has both great advantages and disadvantages. The major disadvantage is that the resynthesis of the stimuli creates artifacts and does not always sound perfectly natural. That might have influenced the rating of the experiment participants. On the other hand, this method offers the possibility to exclude some variables, which would not be achievable with natural stimuli.

Further studies should focus on several possible influences, like the speaker, the gender of the raters, their education or the influence of phonetic knowledge. On top of that, similar research could try to include larger samples, more speakers, different cultural backgrounds of the speakers and raters, and spontaneous instead of laboratory speech. Furthermore, the inclusion of other phonetic aspects into a controlled study should be attempted.

## Acknowledgments

Thank you to Tuarik Buanzur for recording the speakers for the study, Evelin Graupe for helping with the F0 range manipulations, Matthias Hoffmann for the support with the rating profiles, Nele Kiupel for her helpful comments, and – of course – the two speakers and the raters who volunteered for the project.

## References

- [1] Potts, J.: A history of charisma. Palgrave Macmillan, Houndmills, 2009
- [2] Klein, K. J., and House, R. J.: On fire: Charismatic leadership and levels of analysis. *Leadership Quarterly* 6/2 (1995), 183-198
- [3] Antonakis, J., Bastardo, N., Jacquart, P., and Shamir, B.: Charisma: An ill-defined and ill-measured gift. *Annual Review of Organizational Psychology and Organizational Behavior* 3 (2016), 293-319
- [4] Holladay, S. J., and Coombs, W. T.: Communicating visions: An exploration of the role of delivery in the creation of leader charisma. *Management Communication Quarterly* 6/4 (1993), 405-427
- [5] Niebuhr, O., Voße, J., and Brem, A.: What makes a charismatic speaker? A computer-based acoustic-prosodic analysis of Steve Jobs tone of voice. *Computers in Human Behavior* 64 (2016), 366-382
- [6] Rosenberg, A., and Hirschberg, J.: Acoustic/Prosodic and Lexical Correlates of Charismatic Speech. *Proceedings of Interspeech 2005*, 513-516
- [7] Rosenberg, A., and Hirschberg, J.: Charisma perception from text and speech. *Speech Communication* 51 (2009), 640-655
- [8] Biadys, F., Hirschberg, J., Rosenberg, A., and Dakka, W.: Comparing American and Palestinian perceptions of charisma using acoustic-prosodic and lexical analysis. *Proceedings of Interspeech (2007)*, 2221-2224
- [9] Niebuhr, O., Brem, A., and Nowak-Tót, E.: Prosodic constructions of charisma in business speeches – A contrastive acoustic analysis of Steve Jobs and Mark Zuckerberg. *Proceedings of Speech Prosody 8 (2016)*, 1-2
- [10] Berger, S., Marquard, C., and Niebuhr, O.: INSPECTing read speech – How different typefaces affect speech prosody. *Proceedings of Speech Prosody 8 (2016)*, 513–517
- [11] Charpentier, F., and Stella, M.: Diphone synthesis using an overlap-add technique for speech waveforms concatenation. *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP 11 (1986)*, 2015-2018
- [12] Boersma, P., and Weenink, D.: Praat: doing phonetics by computer (Version 6.0.17) (2016), URL: <http://www.praat.org/>
- [13] Leiner, D. J.: SoSci Survey (Version 2.6.00-i) (2014), URL: <https://www.sosicisurvey.de>.
- [14] RStudio Team: RStudio: Integrated Development for R (Version 1.0.44) (2016), URL: <http://www.rstudio.com/>
- [15] Quené, H., Boomsma, G., and Van Erning, R.: Attractiveness of male speakers: effects of voice pitch and of speech tempo. *Proceedings of Speech Prosody 8 (2016)*, 1086-1089