"SPACE FIGHTERS" ON STAGE - HOW THE F1 AND F2 VOWEL-SPACE DIMENSIONS CONTRIBUTE TO PERCEIVED SPEAKER CHARISMA

Oliver Niebuhr[1]

[1]*Centre for Industrial Electronics, University of Southern Denmark, Sonderborg/DK*
oniebuhr@mci.sdu.dk

Abstract: Two perception experiments are carried out, showing that vowel pronunciation matters for perceived speaker charisma. Larger acoustic vowel spaces make speakers sound more charismatic. However, the F1 and F2 dimensions of the vowel space contribute differently to this perceptual effect. The F1 range is related emotional aspects of charisma, such as sounding passionate and captivating, whereas the F2 range is related to cognitive aspects of charisma, such as sounding trustworthy and decided. Implications for computer-based tools of speaker training are discussed.

1 Introduction

Charisma is a complex phenomenon of human perception. The sources of this perception lie in the signaling of competence, passion, and self-confidence on the part of the speaker. On the part of the perceiver, competence signals create trust in the speaker's abilities, while passion and self-confidence signals make the perceiver get inspired or motivated, respectively [27]. In consequence, perceivers become followers of the charismatic speaker's ideas, goals and actions - often on a very subtle level that extends into the neuronal processes of brain activity [19,46,48]. Charisma is, therefore, a very powerful tool that not only determines the economic and professional success of individual people [2,11]. If used by machines, the tool of charisma can even make human perceivers choose fruit over chocolate, fill out longer questionnaires, or take longer routes by car to their destination [15,37].

For one thing, charisma perception is triggered by predefined characteristics of the body such as height [17], age [21], and sex [8,21,41]. For another thing, charisma perception is triggered by characteristics that are under the control of the speaker. These additional characteristics are primarily communicative in nature. One type of communication signals is extralinguistic, such as a speaker's clothing [20], haircut, suntan, and jewelry. The other type of communication signals is linguistic. That is, they rely on the form-function relationships of a specific language. These signals are among the strongest charisma triggers, not least because no other form of communication can convey the sources of charisma - competence, passion, and self-confidence - more effectively than language. Language is, without a doubt, also the charisma trigger that has been best researched. This applies to both the verbal means of language, such as metaphors, three-part lists, anecdotes, contrasts and rhetorical questions (see the Charismatic Leadership Tactics of Antonakis et al. [1]) and to the non-verbal means of language - speech prosody in particular. Whether words or prosodies are the more powerful charisma triggers has not yet been conclusively shown. There are some arguments against words, at least in their role as content-constituting elements [10]. Words as stylistic devices are a different matter, though, not least because contrasts, three-part lists, rhetorical questions, and many other stylistic devices come with specific prosodic patterns; patterns which, by themselves, can already increase perceived speaker charisma.

Which prosodic patterns increase charisma perception? In recent years, experimental phonetic research has been able to show for political and business speakers alike that, among other things, a higher f0 level (with a simultaneously lowered baseline f0), a larger f0 range, a higher acoustic energy level, a faster speaking rate, a shallower spectral tilt, fewer filled pauses, and a shorter prosodic-phrase duration positively are positively correlated with perceived speech charisma [3,29,33,35,38,45,47]. Rosenberg & Hirschberg [45:648], however, also pointed out the limitations of these correlations: "some of these interactions may be at least potentially U-shaped rather than truly linear". Niebuhr et al. [40] not only determined the gender-specific U-shapes for a total of 16 prosodic charisma parameters in three-year series of perception experiments with almost 500 listeners. They also determined the perceptual weight of each of these parameters for listeners of various West Germanic languages. The result is PICSA™, i.e. the Perception-Integrated Charismatic Speech Analysis[1], and its associated training and evaluation system PASCAL (Prosodic Analysis of Speaker Charisma: Assessment and Learning).

While the prosody of charisma has been intensively studied up to the point that allowed the creation of a successful charisma prediction tool [28] like PICSA™/PASCAL, research on the role of sound segments in charisma perception is far less advanced. Virtually every rhetoric manual urges its readers to “clearly articulate every phrase and word” and that a “good articulation conveys competence and credibility” [31:158] and is, thus, “is imperative to develop charisma” [9:138]; see also [16]. Experimental-phonetic research basically underpins these statements [34], but, so far, leaves many questions unanswered.

For example, Niebuhr & Gonzalez [36] found that Steve Jobs and Mark Zuckerberg differ significantly in their perceived speaker charisma. Jobs sounds considerably more charismatic than Zuckerberg both for original stimuli and for de-lexicalized/anonymized stimuli. This difference coincides with a difference in acoustic vowel space size. Jobs' vowel space is almost 40 % larger than that of Zuckerberg, particularly along the back-front dimension, which is associated with the range of the second formant frequency (F2). The larger the F2 range, the farther apart are a speaker's front and back vowels, see Figure 1. Do these results mean that clear vowel pronunciation and, thus, the acoustic distinctiveness of vowel phonemes is a charisma-relevant factor? And is this solely true for the back-front dimension encoded in the F2 range? Such conclusions would contradict [4:176] who explicitly not recommends his charisma learners to work on their vowel pronunciation: “Make sure all the consonants are clear when you are speaking (all the letters that are not A, E, I, O or U) ”. On the other hand, [3] did find an influence of vowel formants on the perception of charisma-relevant speaker attributes, but only for the first formant frequency (F1), which encodes the open-closed dimension of the acoustic vowel space. The larger the F1 range, the farther apart are a speaker's open and closed vowels, see Figure 1.

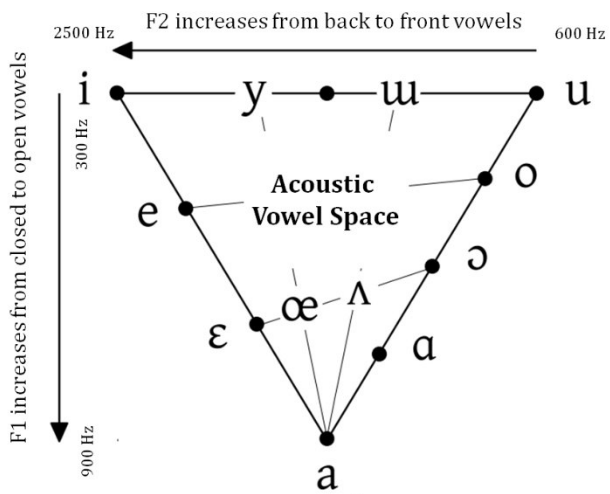

Figure 1 - Acoustic vowel space and its F1/F2 dimensions of open-closed and back-front vowels.

Due to this inconsistent and largely incomplete empirical picture, this paper asks the following questions: Is there a systematic relationship between the vowel-formant distances of a

[1] https://www.allgoodspeakers.com/picsa

speaker or the resulting shape and size of his/her the acoustic vowel space on the one hand and the perception of speaker charisma on the other? If so, is this systematic relationship based on the open-closed dimension of the vowel space, i.e. on the F1 range, and/or on the back-front dimension of the vowel space, i.e. on the F2 range? That is, is it worth for speakers to be "space fighters" on stage, targeting larger acoustic vowel spaces?

These questions are addressed here in two complementary perception experiments. Experiment 1 is based entirely on natural speech data and looks for correlations between the vowel-formant distances (F1 and F2 ranges) of speakers and their charismatic impact on listeners. Natural speech data have a high ecological validity. However, they also mean making between-speaker comparisons; and although the Experiment-1 stimuli were homogenized in many charisma-relevant prosodic parameters prior to being rated by listeners, it cannot be entirely ruled out that voice qualities and other extra-linguistic factors influence the charisma rating of listeners. Therefore, the results of Experiment 1 are re-examined in Experiment 2. Experiment 2 is based on artificially created within-speaker vocal-space manipulations. Thus, effects of the F1 and F2 ranges on perceived speaker charisma can be tested separately and largely without confounding factors, but at the cost of a lower level of ecological validity.

2 Method

2.1 Speech material

Sixteen readings of the Rainbow Passage [14] provided the point of departure for the stimulus creation of Experiments 1 and 2. The Rainbow Passage is one of the most elicited texts within the speech sciences. It was originally developed as a standardized reading exercise for the diagnosis and treatment of speech pathologies. To that end, it was designed such that it covers, in a comprehensive and phonotactically balanced way, the consonant and vowel phonemes of English as well as selected common and rare phoneme sequences. It is due to the latter that the Rainbow passage has also become a popular pronunciation training tool for actors, second-language instructors, and rhetorical coaches [49].

The 16 readers of the Rainbow Passage were native speakers of German (8m/9f), between 20-30 years old, non-smokers, and with no known speaking or hearing disorder. They were recruited from the student pool at the University of Southern Denmark (SDU) and selected to have a similarly strong command of English, i.e. levels C1 or C2 according to the European CEFR scale [22]. The students were recorded in the sound-treated booth of the Acoustics Lab of the Centre for Industrial Electronics at SDU[2], using a AKG C214 professional large-diaphragm condenser studio microphone connected to a Zoom H6 recorder. Recordings were made digitally at 48 kHz sampling rate and 24-bit quantization.

The readers were instructed to produce the Rainbow passage in a fluent, clear way and with an animated tone of voice, like a story-teller in an audio book. All readers were given five minutes to familiarize themselves with the text. Then, they had three rounds of recording, after which they selected their subjectively best performance. The latter was saved as an uncompressed sound file. The readers were informed that their selected sound files would be used as stimuli in a speech-perception experiment. However, neither before nor after the reading task were they informed about the actual topic of this experiment, i.e. the articulatory and acoustic distinctiveness of the readers' vowel tokens and hence the shape and size of their acoustic vowel spaces in terms of F1/F2-ranges. Accordingly, short de-briefing interviews with the readers after the recording showed that none of them guessed the actual purpose of the recording. Most readers assumed to have participated in a voice-attractiveness study.

[2] https://www.sdu.dk/en/om_sdu/institutter_centre/mci_mads_clausen/laboratorier/acoustics+lab

2.2 Stimuli of Experiment 1

The following section of five sentences was extracted from the 16 Rainbow Passage readings: "(1) Throughout the centuries people have explained the rainbow in various ways. (2) Some have accepted it as a miracle without physical explanation. (3) To the Hebrews it was a token that there would be no more universal floods. (4) The Greeks used to imagine that it was a sign from the gods to foretell war or heavy rain. (5) The Norsemen considered the rainbow as a bridge over which the gods passed from earth to their home in the sky." The total duration of the extracted section was about 30 seconds. Each section represented a base stimulus.

The acoustic vowel-space sizes of all 16 base stimuli were analyzed with reference to the four (phonologically long) landmark qualities at the edges of the English phonemic vowel space [44], i.e. (i) the front closed vowel /i/ as in "Greeks", (ii) the back closed vowel /u/ as in "through", (iii) the front open vowel /æ/ as in "imagine", and (iv) the back open vowel /ɑ/ as "passed". The first two formants - F1 and F2 - were automatically extracted for these four landmark qualities using the DARLA system [43]. Then the average acoustic distances (in Hz) between back and front vowels as well as between open and closed vowels were calculated per reader. In other words, the ranges of F1 and F2 were determined per reader.

As expected, we found large differences between the F1/F2 ranges across the 16 readers. Figures 2(a)-(d) show examples of readers whose vowel spaces are overall larger or smaller, or compressed along either the vertical (open-closed) or the horizontal (back-front) axis.

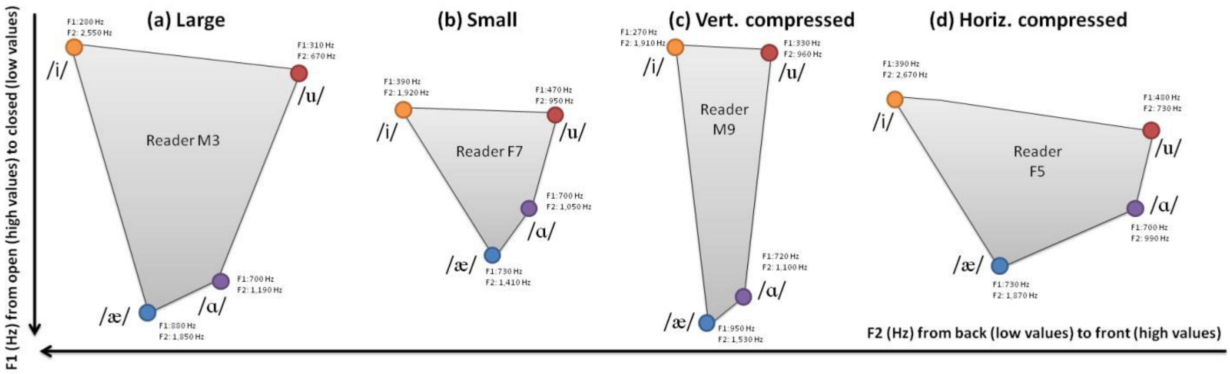

Figure 2 - The F1/F2 values for /i, æ, ɑ, u/ of readers M3, F7, M9, and F5 as examples of (a) large, (b) small, (c) vertically compressed, and (d) horizontally compressed acoustic vowel spaces.

In a following step, the 16 base stimuli were finalized to the actual 16 experimental stimuli by homogenizing the base stimuli with respect to confounding prosodic factors of charisma perception. To that end, the grand means of speaking rate, pitch level, pitch range, and intensity level were determined per gender group, i.e. across the nine female and eight male readers. Then, all base stimuli of that gender group were manipulated such that they became prosodically identical in terms of these grand means. Manipulations were made in a proportional, holistic fashion, using PSOLA resynthesis in PRAAT [7]. Note that there was no need to homogenize further charisma-relevant factors like phrase and pause durations as these factors were inherently similar across all base stimuli thanks to the constant syntactic structure and punctuation in the reading task.

2.3 Stimuli of Experiment 2

A subset of four stimulus pairs was selected from Experiment 1, based on the F1/F2 values of the four landmark vowel qualities (i)-(iv). Each pair consisted of a male and a female stimulus. The differences between the four pairs were as follows. We selected (1) those two (m/f) stimuli with the largest and (2) those two stimuli with the smallest F1 and F2 ranges as well as (3) those two stimuli whose vowel spaces were most strongly vertically compressed (i.e. maximum values for F1 minus F2 range), and (4) those two stimuli whose vowel spaces were most strongly horizontally compressed (i.e. minimum values for F1 minus F2 range).

Now, a DARLA automatic vowel-formant analysis was conducted again, however, not for the four stimulus pairs themselves, but for the corresponding speakers' complete Rainbow Passage readings. Based on the results of this analysis, a further subset of four stimulus pairs was derived from the original subset through manipulation, i.e. in a copy-and-replace procedure. The aim was to create for each of the 4x2 original stimuli a counterpart with an inversely extended or compressed acoustic vowel space. Large formant distances became small ones and small formant differences large ones. Thus, for example, a large vowel space like in Figure 2(a) was turned into a small vowel space like in Figure 2(b) and vice versa. Likewise, a vertically compressed vowel space like in Figure 2(c) was turned into a horizontally compressed vowel space like in Figure 2(d) and vice versa.

To that end, all landmark vowels (i)-(iv) were replaced by identical phonemes but with higher/lower F1/F2 levels, copied from other sentences of the same reader's Rainbow Passage production. In this way, the speakers' vowel spaces were artificially extended or compressed horizontally and/or vertically, using each speaker's own vowel material. To further enhance this extension or compression effect, the replacement went beyond the landmark vowels (i)-(iv) and also included the neighboring short vowel phonemes /ɪ, ʊ, ʌ, ɒ/ as well as the two diphthongs /aɪ/ and /aʊ/, see [44]. Together, the copy-and-replace procedure accounted for almost 75% of all vowels and diphthongs in the five-sentence section of the stimuli (see 2.2). The magnitude of horizontal and vertical vowel-space extension or compression obtained through this copy-and-replace procedure was between 20-30 %. That is, changes in F1 ranges were about 200-250 Hz, and changes in F2 ranges were about 300-400 Hz.

The copy-and-replace procedure was conducted such that disruptions of formant transitions were minimized. That is, the replacements mainly concerned the vowel centers and diphthong onsets. Vowels were artificially lengthened or shortened if necessary with PSOLA resynthesis in order preserve the original rhythmic stress and pitch-accent patterns. Likewise, the f0 patterns of the replaced vowel sections were restored in PSOLA. Nevertheless, the applied copy-and-replace procedure reduced the naturalness of the stimuli to some degree. However, since it were the same vowel phonemes that were replaced at the same places in the text in all stimuli, the presumably negative effect on perceived speaker charisma can be considered constant and, hence, controlled so that the target effects of horizontal and vertical vowel-space compression/extension can still be reliably investigated.

2.4 Listeners

Like the 16 readers, the listeners who took part in Experiments 1 and 2 were also native speakers of German in their twenties, who studied at SDU and did not suffer from any speaking or hearing disorder. Experiment 1 involved 25 listeners, 12 women and 13 men. Experiment 2 was done by 33 listeners, 15 women and 18 men. Listeners received an allowance of 75,- DKK for their participation.

2.5 Procedure

Experiments 1 and 2 were conducted by means of PRAAT MFC [6]. The 16 stimuli were played from a silent tablet computer (no fan, no mechanical HDD) in individually randomized orders and rated with respect to five 10-point scales. They ranged from "not applicable/absent" (0) to "fully applicable" (10). Each scale addressed a different speaker attribute. One scale directly asked to what degree the speaker is 'charismatic'. The remaining four represent a subset of those scales that were successfully used in previous experiments on charisma perception [39]. That is, it is known from these previous studies that the four additionally selected scales are sufficiently sensitive to detect phonetic effects on perceived speaker charisma and, at the same time, sufficiently uncorrelated to effectively cover different aspects of this complex perceptual phenomenon, cf. [28]. The four additional scales were: 'captivating', 'passionate', 'trustworthy' (meaning "capable of living up to his/her promises"), and 'decided'. Note that 'captivating' and 'passionate' relate more to the speaker's emotional state and its charismatic effect, whereas 'trustworthy' and 'decided' relate more to the speaker's cognitive state and its charismatic effect.

Listeners conducted the experiment in individual sessions that took place in the Acoustics Lab of the Centre for Industrial Electronics at SDU. Each session started with watching an e-learning video (4:09 minutes) about the concept, origin, and relevance of speaker charisma [27]. The video was followed by the instruction to listen to 16 speakers who present, consecutively, the same 30-second excerpt of a popular English short story called "The Rainbow Passage". Participants were asked to ignore the verbal content of the presentations for their rating task. Instead they were asked to focus on the speakers and their presentation performances and to rate, on this basis, how charismatic each speaker had sounded. Listeners were informed that they would receive five difference scales to conduct this rating task, and that the ratings were to be made after the end of each speaker's presentation. In fact, the PRAAT-MFC script was written such that it was not possible for participants to start the rating task before the end of a stimulus [6].

After the instruction, participants received three familiarization trials to practice the rating task. The stimuli of these familiarization trials came from readings of the "Rainbow Passage" excerpts (see 2.2) by the first author (ON) and two (female) co-workers (JV, KD).

Participants listened to the stimuli via Bose QuietComfort 35 headphones that were, with active noise cancellation switched on, directly connected to the tablet PC through a 3.5 mm jack plug. A complete experimental session from the initial e-learning video presentation through all stimulus ratings to the final de-briefing took about 30 minutes.

3 Results

3.1 Results of Experiment 1

Correlation tests (Persons Product Moment Correlations, PMCC) were conducted between a reader's mean F1 and F2 ranges on the one hand and his/her assessment on the five charisma rating scales on the other. Thus, ten PMCC tests were conducted. One set of five tests addressed the correlations between a reader's F1 ranges and the charisma rating scales. The other set of five tests did the same, but for a reader's F2 ranges. Each correlation analysis was based on a total of 400 values (16 readers x 25 listener assessments). The Benjamini-Hochberg method was used to correct alpha-error/p-levels for multiple testing.

Table 1 - Results summary of the Pearson Product-Moment Correlations for the data of 16 readers and 25 listeners (N=400) per rating scale. Stats columns show r, p, and , η^2.

Scale type	Scale attribute	Stats F1 range (*df*=398)	Stats F2 range (*df*=398)
charisma direct	charismatic	$r=0.53$, $p<.0001$, $\eta^2=.28$	$r=0.48$, $p<.0001$, $\eta^2=.23$
emotion-related	captivating	$r=0.56$, $p<.0001$, $\eta^2=.31$	$r=0.40$, $p<.0001$, $\eta^2=.16$
emotion-related	passionate	$r=0.62$, $p<.0001$, $\eta^2=.38$	$r=0.46$, $p<.0001$, $\eta^2=.22$
cognition-related	trustworthy (capable)	$r=0.39$, $p<.0001$, $\eta^2=.15$	$r=0.71$, $p<.0001$, $\eta^2=.51$
cognition-related	decided	$r=0.37$, $p<.0001$, $\eta^2=.14$	$r=0.55$, $p<.0001\eta^2=.30$

All ten PMCC tests yielded highly significant results. Correlations were positive, i.e. the greater a speaker's F1 range and/or F2 range the more charismatic as well as captivating, passionate, trustworthy (meaning "capable of living up to his/her promises"), and decided did s/he sound in the ears of listeners. Table 1 summarizes the PMCC test statistics. Note that the F1 ranges were most strongly linked to sounding passionate and captivating, whereas differences in the F2 ranges most strongly correlated with a speaker's perceived trustworthiness and decidedness. The term charisma itself fell in between these two levels of correlation. That is, the degree to which a speaker sounded charismatic strongly relied on both the F1 range and the F2 range, slightly more on the former than on the latter, though.

Figures 3(a)-(b) illustrate the PMCC results for the two strongest correlations: F1 range and perceived passion as well as F2 range and perceived trustworthiness. The x-axes show the mean F1 or F2 ranges of the 16 readers in an ascending order (left to right). The y-axes show the ratings of the 25 listeners. The dots in each figure represent the mean ratings, the vertical bars represent the rating span, i.e. highest and lowest ratings across all 25 listener.

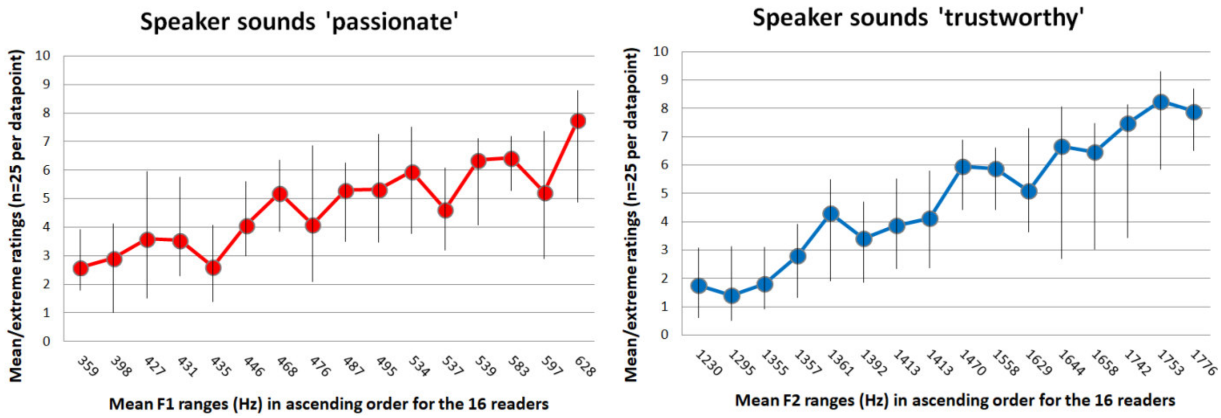

Figure 3 - The 16 readers' mean F1 ranges (a, left) and mean F2 ranges (b, right) and their positive correlations with perceived passion (r=0.62) and trustworthiness (r=0.71); N=400 per correlation. Dots show the mean ratings and vertical bars the range of ratings across the 25 listeners.

3.2 Results of Experiment 2

The results of Experiment 2 were statistically analyzed with RM (repeated-measures) ANOVAs. Separate RM-ANOVAs were run for each of the five rating scales, with 'listener rating' (0-10, N=33) as the dependent variable. Each RM-ANOVA was based on the fixed factors Vowel Space Size (4 levels: large, small, vertically compressed, horizontally compressed) and Vowel Space Manipulation (2 levels: original, inverted). Additionally, making use of the controlled, gender-balanced conditions (unlike in Exp.1) and taking into account the results of [8,21,39], Reader Gender (m/f) was included as a third fixed factor in the RM-ANOVAs. Listener (33 levels) was included as a covariate.

Table 2 - Summary of the test statistics of the five RM-ANOVAs. As all five analyses yielded similar results, only the minimum and maximum values across the five analyses are shown here.

	df	F	p	η_p^2
Vowel Space Size	3,93	113.965 - 177.762	<0.001	0.719 - 0.895
Vowel Space Manip.	1,31	7.636 - 14.414	<0.001 - <0.0001	0.198 - 0.333
Interaction Size*Manip	3,93	14.813 - 20.023	<0.001 - <0.0001	0.373 - 0.448
Reader Gender	1,31	5.851 - 12.946	<0.5 - <0.01	0.164 - 0.252

All five RM-ANOVAs yielded significant main effects of both Vowel Space Size and Vowel Space Manipulation as well as of Reader Gender, see the summary in Table 2. Moreover, the results of all RM-ANOVAs were similar insofar as Reader Gender was always the weakest main effect and in that there were no significant two-way or three-way interactions except for that of Vowel Space Size and Vowel Space Manipulation. The main effect of Reader Gender reflects that ratings were lower for the female than for the male readers. This was found for all five scales. Which rating differences caused the other main effects an interactions is detailed below, based on additional multiple-comparisons tests (with Sidak correction) that were conducted between the levels of all fixed factors.

Recall for the factor Vowel Space Size that 'large' refers to those male and female readers with the largest ranges in F1 *and* F2, whereas 'small' refers to those male and female readers with the smallest ranges in F1 *and* F2. In contrast, 'vertically compressed' means that the F1 range is relatively large, while, at the same time, the F2 range is relatively small. The opposite F1-F2 relation holds for the 'horizontally compressed' condition. Against this background, Figures 4(a)-(c) illustrate, for three of the five rating scales, the typical results patterns that underlie the significant effects in all RM-ANOVAs.

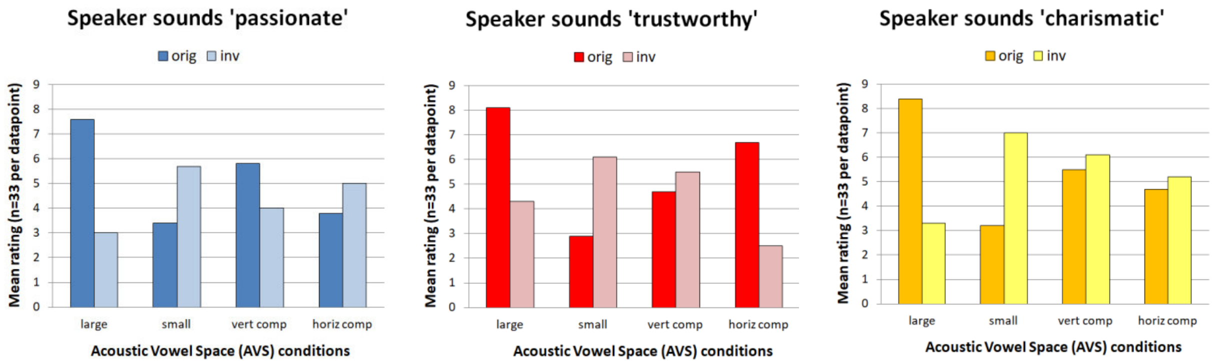

Figure 4 - Representative results pattern for the four Vowel Space Size conditions of Experiment 2. of Three out of five rating scales are shown. Each bar represents the mean rating of 33 listeners for two stimuli (m/f ratings are pooled).

First, the 'large' Vowel Space Size condition resulted in the highest ratings of charisma and all related attributes like passion and trustworthiness (Fig.4). Second, the 'small' Vowel Space Size condition resulted in the lowest ratings of charisma and all related attributes like passion and trustworthiness. Third, compared to these two more symmetrical vowel space conditions, the two conditions with strongly asymmetrical vowel spaces, i.e. 'vertically/horizontally compressed', yielded intermediate ratings levels on all scales, including 'passionate' and 'trustworthy' as well as 'charismatic' itself. That is, listener ratings were significantly lower than in the 'large' vowel-space condition, but significantly higher than in the 'small' vowel-space condition. Fourth, for the two emotion-related scales 'passionate' and 'captivating' it was the horizontal compression (i.e. a small F1 range) that more severely lowered ratings to a level significantly below that of the vertical compression. The opposite applied to the two cognition-related scales 'trustworthy' and 'decided'. Here it was the vertical compression (i.e. a small F2 range) that more severely lowered ratings to a level significantly below that of the horizontal compression (together, this constitutes the reason for the interaction Vowel Space Size * Vowel Space Manipulation). Fifth, in the context of this interaction, the variation of ratings along the charisma scale itself was more similar to the two emotion-related than to the two cognition-related scales. Finally, sixth, the artificial inversion of the vowel space sizes through copying and replacing vowels (e.g., turning 'large' into 'small', 'vertically compressed' into 'horizontally compressed' and vice versa) indeed significantly shifted listener ratings in the direction of the corresponding original condition. That is, for example, an artificially created large vowel space caused similar rating shifts relative to the other conditions as its original, naturally produced large vowel space counterpart. However, none of these artificially created rating shifts reached the magnitude of caused by its original, naturally produced counterpart, which is most likely due to artifacts of the copy-and-replace procedure.

4 Discussion

In rhetorical practice, a clear pronunciation is one of the key recommendations for learners of charismatic public speaking. This recommendation is basically consistent with the results of empirical phonetic research [3,34,28,36]. However, if and how vowel pronunciation makes a contribution to a speaker's charismatic impact on listeners has so far been unclear. For example, while [4] advises his learners to focus on a clear consonant pronunciation, thus questioning the very need to work on vowel pronunciation, findings of phonetic production and perception experiments suggest that vowels also have a noticeable influence on the perception of speaker charisma [3,36].

In order to expand and refine our knowledge on the role of vowels in perceived speaker charisma, the following research questions were asked in this study: Is there a systematic relationship between the vowel-formant distances of a speaker or the resulting shape and size of his/her the acoustic vowel space on the one hand and the perception of speaker charisma on the other? If so, is this systematic relationship based on the open-closed dimension of the vowel space, i.e. on the F1 range, and/or is it rather based on the back-front dimension of the vowel space, i.e. on the F2 range, or do even both formant ranges play a role?

By means of two perception experiments, this paper was able to provide answers to these questions. Firstly, it was shown in Experiment 1 that not only consonants matter. *Vowels matter as well*. Their pronunciation is important for a speaker's charismatic impact on listeners. Larger acoustic vowel spaces and, thus, larger vowel distances in terms of F1 and F2 ranges lead to an increased speaker charisma. This matches with the expected relationship between speaker charisma and pronunciation. The clearer the pronunciation (of vowels) and, thus, the more the vowels are produced acoustically distinct from one another, the more charismatic is the speaker.

The second key finding of the present paper concerns the shape of the vowel space in relation to speaker charisma. While the findings of [36] on a more charismatic speaker (Steve Jobs) and a less charismatic speaker (Mark Zuckerberg) suggest that it is primarily the F2 range the vowel space that determines speaker charisma, [3] only find a correlation between F1 and charisma-relevant speaker attributes. Experiment 2, in particular, clarifies this issue, based on within-speaker manipulations of the vowel space shape and size: Both the F1 range *and* the F2 range influence perceived speaker charisma. However, the quality of this influence is *formant-specific*. The F1 range primarily determines how passionate and captivating a speaker sounds. This means that the F1 range primarily correlates with the emotional aspects of speaker charisma. In contrast, differences in the F2 range make a speaker sound more or less trustworthy and decided. Thus, the F2 range correlates primarily with the cognitive and capability-related aspects of speaker charisma. That the rating scale on charisma itself correlated more with variation along the F1 range suggests that charisma is, overall, more strongly shaped by emotional aspects (see Emotional Contagion [19]) than by aspects of competence or capability. This also fits in with the weighting of the emotion-related prosodic parameters (e.g., f0 range) relative to that of the capability-related prosodic parameters (e.g., phrase duration) in the charisma assessment algorithm PICSA™ and its dynamic scoring tool PASCAL [40].

Beyond the two key findings, the present paper also replicated the results of [8,21,39] again. That is, all else equal, women have an about 20 % lower charisma effect on listeners than men. In Experiment 2, this gender-specific charisma bias occurred in the form of a significant main effect of Reader Gender that was found for all five charisma-related rating scales. That is, compared to the male readers, the female readers were rated less 'passionate', less 'decided', less 'trustworthy', less 'decided' as well as, overall, less 'charismatic'. Niebuhr et al. [40] discuss the origins of this phenomenon and describe how women can reduce or overcome this bias through targeted tone-of-voice training.

The most significant implication of the present findings is that *vowel pronunciation must not be ignored* in rhetorical speaker charisma training. Of course, further studies are needed to quantify the relevance of vowel pronunciation for speaker charisma in relation to the effect sizes of prosodic features and consonant pronunciation. The present findings do not allow such a relative impact quantification, but they contain indications. One indication are the robust correlations that have emerged between the vowel-formant ranges on the one hand and perceived charisma and its four related attributes on the other, even for between-speaker stimuli. Another indication is that these between-speaker stimuli yielded ratings that varied by factor 3-4 across the stimulus set. For example, the slope factors of the linear correlations shown in Figures 3(a)-(b) are 0.0177 and 0.0119 respectively. That is, it takes less than 100 Hz of a difference in a speaker's F1 or F2 range to change his/her perception as being 'charismatic', 'passionate', 'trustworthy' etc. by one scale point on a ten-point scale. This is a fairly steep correlation slope and, thus, a fairly powerful effect.

Together, this suggests a relatively high relevance of vowel pronunciation for perceived speaker charisma. Such a paramount importance of vowels for speaker charisma would also make sense from the point of view of vowels as key players in speech communication. Vowel literally act as pivots in speech production, for example, with regard to the coordination of articulatory gestures [42] and the control, timing, and perception of prosodic elements [5,24,32]. Furthermore, vowels are perceptually the most prominent elements in speech due to their intrinsically high levels of acoustic energy and duration [30].

Unlike consonants, however, vowels are not easy to train. They lack tactile feedback and a biunique link between articulation and acoustics [25]. There are computer-aided systems for real-time vowel-formant visualization that are successfully used in foreign language teaching and pronunciation training for the deaf [13,23,26]. In future research, it is a goal of the author and his team to further develop such a system for charisma training, based on PICSA and PASCAL. Existing systems cannot simply be adopted, for example, because a charisma-oriented vowel-pronunciation training must be done in connected-speech contexts rather than on the basis of isolated vowel sounds. Another reason is that the feedback given to speakers must focus on acoustic distinctivity rather than on phonological distinctivity. That is, the training would not be about distinguishing /i/ from /u/ or /i/ from /æ/. It would be about maximizing these phonological distinctions at the level of phoneme-internal pronunciation variants. In addition, gender-specific and stylistic formant differences need more attention than in foreign-language training and pronunciation training for the deaf.

Finally, it is also important in this R&D context to go beyond a central limitation of the present study: the use of L2 English speakers. Initial results of follow-up experiments are promising in that they show that the results presented here also apply to L1 speakers of German and English. Next, the role of vowel-induced charisma will be tested for Danish, a language that is a worthwhile research object not least because of its high number of vowel phonemes (20 and more, [18]). Furthermore, it is assumed, further research pending, that vowel pronunciation training with the aim to enlarge F1 and/or F2 ranges has a beneficial side effect on a speaker's acoustically projected body size [50], i.e. on how tall or big a speaker sounds in the ears of listeners. Given that this body-size factor matters for perceived charisma [17], it seems not only possible but also desirable to try to increase a speaker's acoustically-projected body size through vowel-based pronunciation training. In any event, the ongoing R&D work on vowel pronunciation will lend a new dynamic to the understanding and training of speaker charisma. What is clear at this early stage already is that it is worth for a public speaker to be a "space fighter" on stage, fighting against physiological constraints and the biological imperative of energy-saving hypo-articulation [12] for a larger, more charismatic acoustic vowel space.

5 References

[1] ANTONAKIS, J., M. FENLEY, M., S. LIECHTI: Can charisma be taught? Tests of two interventions. *Acad. Manag. Learn. Educ. 10*, 374–396, 2011.

[2] ANTONAKIS, J., G. D'ADDA, R. Weber, C. ZEHNDER: Just words? Just speeches? On the economic value of charismatic leadership. *NBER Rep. 4*, 2014.

[3] APPEL, J., C. PESCHER: Formanten und ihr Einfluss auf den wahrgenommenen Status von Präsentierenden aus Sicht der Zuhörer. *Sprechen 54*, 4-8, 2012.

[4] BARKER, A: *Improve Your Communication Skills*. London: Replika Press, 2011.

[5] BARNES, J., N. VEILLEUX, A. BRUGOS, S. SHATTUCK-HUFNAGEL: Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology 3*, 337-383, 2012.

[6] BECK, J.: ExperimentMFC - Erstellung und Auswertung eines Perzeptionsexperimentes in Praat. *KALIPHO 3,* 81-113, 2015.

[7] BOERSMA, P.: Praat, a system for doing phonetics by computer. *Glot international 5*, 341-345, 2002.

[8] BROOKS, A.W., L. HUANG, S.W. KEARNEY, F.E. MURRAY: Investors prefer entrepreneurial ventures pitched by attractive men. *Proc. National Academy of Sciences of the United States of America (PNAS) 111*, 4427-4431, 2014.

[9] CAMPER BULL, R: *Moving from Project Management to Project Leadership: A practical guide to leading groups*. Boca Raton: CRC, 2010.

[10] CASPI, A., R. BOGLER, O. TZUMAN: "Judging a Book by Its Cover": The Dominance of Delivery Over Content When Perceiving Charisma. *Group & Organization Management.* https://doi. org/10.1177/1059601119835982.

[11] DAVIS, B.C, K.M. HMIELESKI, J.W. WEBB, J.E. COOMBS: Funders' positive affective reactions to entrepreneurs' crowd-funding pitches: The influence of perceived product creativity and entrepreneurial passion. *J. of Bus. Vent. 32*, 90–106, 2017.

[12] DOCHERTY, G.J.: Speech Production and Speech Modelling. *Journal of Phonetics 20*, 403-410, 1992.

[13] DOWD, A., J.R. SMITH, J. WOLFE: Learning to pronounce vowel sounds in a foreign language using acoustic measurements of the vocal tract as feedback in real time. *Language and Speech 41*, 1-20, 1998.

[14] FAIRBANKS, G: *Voice and articulation drillbook*, 2nd edn. New York: Harper & Row, 1960.

[15] FISCHER, K., O. NIEBUHR, L.C. JENSEN, L. BODENHAGEN: Speech Melody Matters — How Robots Profit from Using Charismatic Speech. *ACM Transactions on Human-Robot Interaction 9*, 1-21, 2019.

[16] FRESE, M., S. BEIMEL, S. SCHOENBORN: Action training for charismatic leadership: Two evaluations of studies of a commercial training module on inspirational communication of a vision. *Personnel Psychology 56*, 671–697, 2003.

[17] GRABO, A., B.R. SPISAK, M. van VUGT: Charisma as signal: An evolutionary perspective on charismatic leadership. *The Leadership Quarterly 28*, 473-485, 2017.

[18] GRØNNUM, N.: Illustrations of the IPA: Danish. *Journal of the International Phonetic Association 28*, 99-105, 1998.

[19] HATFIELD, E., J.T. CACIOPPO, R.L. RAPSON: Emotional contagion. *Current directions in psychological science 2*, 96-100, 1993.

[20] HOLMAN, R.H.: Clothing As Communication: an Empirical Investigation. *Advances in Consumer Research 7*, 372-377, 1980.

[21] JOKISCH, O., V. IAROSHENKO, M. MARUSCHKE, H. DING: Influence of Age, Gender and Sample Duration on the Charisma Assessment of German Speakers. *Proc. 29th Conference on Electronic Speech Signal Process, Ulm, Germany*, 1-8, 2018.

[22] JONES, N., N. SAVILLE: European language policy: Assessment, learning, and the CEFR. *Annual Review of Applied Linguistics 29*, 51-63, 2009.
[23] KRÖGER, B.J., V. Graf-BORTSCHELLER, A. LOWIT: Two- and three-dimensional visual articulatory models for pronunciation training and for treatment of speech disorders. *Proc. Interspeech 2008, Brisbane, Australia*, 2639–2642, 2008.
[24] LADD, D.R.: Segmental anchoring of pitch movements: Autosegmental association or gestural coordination? *Italian Journal of Linguistics 18*, 19, 2006.
[25] LADEFOGED, P.: *Three areas of experimental phonetics*. London: Oxford Univer. Press, 1967.
[26] MASSARO, D.W., Y. LIU, T.H. CHEN, C. PERFETTI: A multilingual embodied conversational agent for tutoring speech and language learning. *Proc. Interspeech 2006, Pittsburgh, USA*, 825–828, 2006.
[27] MICHALSKY, J.: Einfach überzeugend! Mehr Charisma durch das Verständnis von Sprache und Stimme. *Open Course, Virtuelle Hochschule Bayern*, 2019.
[28] MICHALSKY, J., O. NIEBUHR: Myth busted? Challenging what we think we know about charismatic speech. *Acta Universitatis Carolinae Philologica 2019/2*, 27-56, 2019.
[29] MIXDORFF, H., O. NIEBUHR, A. HÖNEMANN: Model-based prosodic analysis of charismatic speech. *Proc. 9th International Conference of Speech Prosody, Poznan, Poland*, 1-5, 2018.
[30] MÖBIUS, B.: Gestalt Psychology meets phonetics–An early experimental study of intrinsic F0 and intensity. *Proc. 15th International Congress of Phonetic Sciences, Barcelona, Spain,* 2677-2680, 2003.
[31] MORTENSEN, K.W: *The Laws of Charisma: How to Captivate, Inspire, and Influence for Maximum Success*. New York: Amacom, 2011.
[32] NIEBUHR, O.: The acoustic complexity of intonation. *Nordic prosody XI*, 15-29, 2013.
[33] NIEBUHR, O., J. VOßE, A. BREM: What makes a charismatic speaker? A computer-based acoustic prosodic analysis of Steve Jobs tone of voice. *Computers and Human Behavior 64*, 366–382, 2016
[34] NIEBUHR, O.: Clear Speech – Mere Speech? How segmental and prosodic speech reduction shape the impression that speakers create on listeners. *Proc. Interspeech 2017, Stockholm, Sweden*, 894–898, 2017.
[35] NIEBUHR, O., R. SKARNITZL, L. TYLEČKOVÁ: The acoustic fingerprint of a charismatic voice – Initial evidence from correlations between long-term spectral features and listener ratings. *Proc. 9th International Conference of Speech Prosody, Poznan, Poland*, 359-363, 2018.
[36] NIEBUHR, O., S. GONZALEZ: 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 24*, 343–355, 2019.
[37] NIEBUHR, O., 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. *Lecture Notes in Computer Science 11620*, 375–390, 2019.
[38] NIEBUHR, O., K. FISCHER: Do not hesitate! – Unless you do it shortly or nasally: How the phonetics of filled pauses determine their subjective frequency and perceived speaker performance. *Proc. Interspeech 2019, Graz, Austria*, 1-5, 2019.
[39] NIEBUHR, O., S. WRZESZCZ: 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. *Proc. 13th International Pragmatics Association Conference, Hong Kong, China,* 1-2, 2019.
[40] NIEBUHR, O., S. TEGTMEIER, T. SCHWEISFURTH: Female speakers benefit more than male speakers from prosodic charisma training – A before-after analysis of 12-weeks and 4-h courses. *Frontiers in Communication 4/12*, 2019.

[41] NOVÁK-TÓT, E., O. NIEBUHR, A. CHEN: A gender bias in the acoustic-melodic features of charismatic speech? *Proc. 18th International Interspeech Conference, Stockholm, Sweden*, 2248-2252, 2017.
[42] ÖHMAN, S.: Numerical models of coarticulation. J. Acoust. Soc. Am 40, 310-320, 1967.
[43] REDDY, S., J.N. STANFORD: Toward completely automated vowel extraction: Introducing DARLA. *Linguistics Vanguard 1*, 15-28, 2015.
[44] ROACH, P.: British english: Received pronunciation. *Journal of the International Phonetic Association 34*, 239-245, 2004.
[45] ROSENBERG, A., J. HIRSCHBERG: Charisma perception from text and speech. *Speech Communication 51*, 640–655, 2009.
[46] SCHJOEDT, H., A.W. STODKILDE-JORGENSEN, T.E. GEERTZ, A. LUND: The power of charisma-perceived charisma inhibits the frontal executive network of believers in intercessory prayer social cognitive and affective. *Neuroscience 6*, 119–127, 2010.
[47] SIGNORELLO, R., F. D'ERRICO, I. POGGI, D. DEMOLIN: How Charisma Is Perceived from Speech: A Multidimensional Approach. Privacy, Security, Risk and Trust (PASSAT). *International Conference on Social Computing (SocialCom), Amsterdam, The Netherlands,* 435-440, 2012.
[48] SOORJOO, M.: *Here's the Pitch: How to Pitch Your Business to Anyone, Get Funded, and Win Clients.* Hoboken: John Wiley & Sons, 2012.
[49] VUKOVIC, J., B.C. JONES, L. DEBRUINE, D.R. FEINBERG, F.G. SMITH, A.C. LITTLE..., J. MAIN: Women's own voice pitch predicts their preferences for masculinity in men's voices. *Behavioral Ecology 21,* 767-772, 2010.
[50] XU, Y., A. LEE, W.L WU, X. LIU, P. BIRKHOLZ: Human vocal attractiveness as signaled by body size projection. *PloS one 8, e62397*, 2013.