Interspeech 2024
1-5 September 2024, Kos, Greece

How rhythm metrics are linked to produced and perceived speaker charisma

Oliver Niebuhr, Nafiseh Taghva

[1]Centre for Industrial Electronics, University of Southern Denmark, Sonderborg, Denmark
[2]Department of Foreign Languages and Linguistics, Shiraz University, Iran
olni@sdu.dk, taghvanafiseh@gmail.com

Abstract

Based on a medium-sized sample of English investor-oriented business-idea presentations (so-called "investor pitches"), the present paper investigates the links between speech rhythm and perceived speaker charisma. Eight trained public speakers are recorded while performing the same investor pitch twice, once in an emotionally-neutral matter-of-fact fashion and once charismatically, i.e. in an expressive, committed onstage presentation style. The recorded presentations were rated by 21 listeners for their degree of perceived speaker charisma – and additionally acoustically analyzed in terms of established duration-based rhythm measures such as ∅, Δ, and PVI. We find significant rhythmic differences between the matter-of-fact and charismatic presentation performances and, in conjunction with the perception results, we show that consonantal rhythmic elements play a bigger role in the perception than in the production of a rhythmic charisma, and that especially the duration variation of larger rhythm elements correlates positively and gender-independently with charisma ratings. The findings are discussed in light of previous studies with their practical implications.

Index Terms: speech rhythm, charisma, prosody, gender.

1. Introduction

How others judge us and the influence we have on others are determined to a remarkably large extent by the way we speak, phonetically and, especially, prosodically. For example, experimental and field studies show that over 70 % of emotion types perceived by listeners can be identified based on prosodic patterns alone [1]. Signs of stress and depression can be derived even more precisely from prosody, as recent meta-analyses show [2,3]. Regarding our influence on others, a study by Chen et al. [4] showed that, when modeling human ratings of public speaking performance, algorithms benefit most from phonetic input, over both visual and lexical input. In a more commercially oriented rating task, Nagano et al. [5] found that prosody significantly determined the willingness to purchase an offered product. Similarly, Amari et al. [6] demonstrated in a machine-learning context in which deep neural networks learned to predict the outcomes of counselor-customer sales interactions (purchase vs. non-purchase) that "acoustic features were more important than linguistic features for this task" (p.8). That is, with the acoustic (mainly prosodic) features alone, the machine-learning algorithm achieved an accuracy of 63-67 %, depending on whether it used the signals of the consultant or the customer.

In our own line of research, we were able to influence through the prosody of both human and machine speakers whether people booked city trip A or B, or whether or not they followed a navigation device's route instructions while driving. We could also predict how well entrepreneurs would perform on investor pitches and students on their oral exams [7,8,9,10].

The special property of individuals that can bring about such changes in opinion and/or behavior is called charisma. If it originates from communication signals, we can also call it perceived speaker charisma. Inspired by the early investigations of political public speeches by Touati [11], Duez [12], Strangert & Gutsafson [13], and Rosenberg & Hirschberg [14], a number of studies have identified and quantified the prosodic correlates of perceived speaker charisma. Features of pitch, tempo, and pause always stood out in such correlations, and loudness or vocal effort also seem to contribute significantly to the predicted speaker effects or performances [15,16], cf. also [17,18].

Crucially, pitch, tempo, pauses, and loudness are not only core dimensions of prosody, they are also the key ingredients of speech rhythm [19]. Given that, it is all the more astonishing that rhythmic correlates of perceived speaker charisma have hardly been a subject of discussion so far. For example, the advice literature on rhetoric either does not address rhythm as a characteristic of charisma at all or confines itself to vague statements like "Charismatic rhetoric uses means such as rhythm, repetition and alliteration" [20:183] or "Great speakers vary rhythm and emphasis to influence their listeners and lead them towards significant portions of their speech" [21:329]. Similarly, in the context of improved strength and effectiveness, [22:79] recommends to "work on developing a pattern or rhythm to your speech," while [23:312] warns his readers of an "overemphasized and choppy rhythm", however, without explaining how such a rhythm sounds like or can be avoided.

Even on a scientific level, the connection between charisma and rhythm is largely unexplored. Niebuhr et al. [24], in their prosodic case study of Steve Jobs' charismatic presentation style, found that his acoustics is "indicative of a speech rhythm whose basic characteristic is diversity and variation between two extremes that are referred to as 'morse-code rhythm' and 'machine-gun rhythm'" (p. 374). To the best of our knowledge, the only study of charismatic speech production and perception specifically devoted to rhythm is that of Bosker [25]. Comparing the presidential debates between Hillary Clinton and Donald Trump in 2016, and based on energy-normalized modulation spectra, Bosker found that Clinton outperformed Trump in terms of a more contrasting and regular syllable rhythm, particularly at frequencies below 4 Hz. That such a more contrasting and regular syllable rhythm is also valued by listeners in the form of significantly higher charisma ratings was corroborated by Bosker in a sub-sequent perception experiment with delexicalized, low-pass filtered speech stimuli.

In summary, there are clear indications of the relevance of speech rhythm for the production and perception of speaker charisma. But, unlike for many other prosodic dimensions, there is currently no precise understanding of the connections between charisma and established rhythm measures, such as ΔC, VarcoV, or the nPVI. The study by [24] uses such rhythm

10.21437/Interspeech.2024-277

measures, but without a comparative or perceptual approach, whereas the study by [25] is both comparative and perceptual, but does not include any of the established rhythm measures.

It is important to note that although measures like ΔC, ΔV and PVI are established, they neither measure rhythm as such nor as a whole. Rather, they capture properties of the sequential structuring of speech. As is already indicated in our reference to [25], rhythm is not just sequential structuring. It is a more or less regular alternation between more or less perceptually salient speech units, such as syllables – and alternation that is, moreover, significantly shaped and controlled by top-down processing and expectations, see [37]. In this respect, we agree with the critical view of, for example, Arvaniti [26] on measures such as ΔC, ΔV and PVI. In order to express this difference and the limits of our approach, we will therefore not speak of rhythm per se in the following, but of rhythm metrics or rhythm measures. Nevertheless, it remains sensible and useful (and original) to test the extent to which such measures are related to perceived speaker charisma.

Our present study is therefore the first to address this research gap. We compare on a within-speaker basis (1) which rhythm parameters change in which direction from neutral to charismatic performance, and (2) how strongly these respective changes correlate with perceived speaker charisma. Established acoustic measures from ΔC to various pairwise variability indices (PVIs) served as our rhythm parameters. Further details about the method are summarized in Section 2 below.

2. Method

2.1. Elicitation of presentations

The elicitation task was done with 8 experienced public speakers, 4 men and 4 women, aged 22-37 years. All were fluent non-native speakers of English and performed the same investor-pitch presentation (in L2 English). TOEFL iBT scores for speaking (cf. [27]) were available for all 8 speakers. Their scores varied between 22 and 28 (∅ 25.2, SD 1.36). The pitch itself, pre-formulated on paper, consisted of 218 words and was written in English in a persuasive style that took into account the verbal Charismatic Leadership Tactics of [28]. The topic was a newly developed mobile app that would help companies track their employees' work-time. The text was adopted an e-learning course on "How to write a killer elevator pitch", [29].

All speakers thoroughly familiarized themselves with the text of the pitch and signaled when they were ready to perform it in an audience-oriented presentation style. The performances were recorded in the sound-attenuated Innovation Lab at the University of Southern Denmark (SDU). The lab is the facility at SDU where investor pitches are held and trained. The speakers were all familiar with the place and with giving talks there. Recordings were made digitally at 44.1 kHz, 24-bit, using a ZOOM H6 recorder. Participants wore a hi-fi headset microphone (Sennheiser HSP 4 EW) for the recording.

Two presentation conditions were consecutively recorded and speakers were instructed accordingly. First, the pitch was to be given in an emotionally neutral, fact-oriented newsreader's style with no special audience in mind. This presentation will henceforth be referred to as the 'neutral' condition. Then, the pitch was to be given in an expressive, committed onstage presentation style, aimed at persuading an imagined jury of potential investors to invest money into the new app. It is henceforth referred to as the 'charismatic' condition. Note that the order of these two performances was not randomized across speakers as, according to our experience, a persuasive presentation has a stronger influence (in terms of a prosodic "afterimage") on a subsequent neutral presentation than vice-versa. In other words, we see in public-speaking training that speakers can often switch more easily and consistently from a neutral to a charismatic style than the other way around.

2.2. Charisma ratings of presentations

The charisma ratings were done by 21 listeners, twelve men and nine women, aged between 29 and 45 years. All listeners were fluent non-native English speakers, just like the speakers in the experiment. They had prior experience in evaluating investor-pitch performances, but neither did they receive training in phonetics or rhetoric nor were they informed about the background and aims of the rating task.

The rating task was conducted at the CIE Acoustics Lab at SDU in in six sessions, with three to five listeners each. Listeners sat at acoustically and visually isolated PC workstations. Stimuli were played over headphones (Bose QuietComfort II) with ANC switched on to further improve sound insulation. A start screen on the PC collected some metadata and explained the concept of charisma to the listeners. After that, the charisma ratings were made per stimulus on a Likert scale from 1 (highest level) to 6 (lowest level). Listeners were instructed to make their ratings only after the stimulus had been fully played. The stimuli themselves were played in randomized order.

As the investor pitches were too long to be included in the rating task in full length, only a two-sentence excerpt was used: "I've come up with an easy way for both employees and employers to lock and keep track of hours using just their cell phones and an app I have designed. It has reduced timecard inconsistencies and paycheck errors by 90%, saving both your time and money." The excerpt was the core statement of the investor pitch and, as suggested by informal listening, also the section that differed most between the neutral and charismatic conditions. Moreover, it was a section that came from about the middle of the pitch so that we were able to avoid prosodic phenomena related to turn boundaries.

2.3. Annotation and rhythmic measures

The presentation recordings were first automatically segmented and annotated at sound-segment, syllable, phrase (pause), and sentence levels using WebMAUS [30] with its built-in rules for English phonology. All automatically set boundaries were manually checked and, if required, adjusted with combined reference to waveform and broadband spectrogram following the criteria of phonetic segmentation summarized in Machač & Skarnitzl [31]. The output of WebMaus was saved per recording as a Praat TextGrid file.

Interval-based speech-rhythm measurements were then derived from these TextGrid files using a Praat script by Volker Dellwo, see [32] for further details. Three established types of measures were calculated based on interval durations: mean duration or distance, standard deviation (Δ), and pairwise variability index (PVI), see [33,34,35]. The three measures were applied to the following intervals: individual vowels and consonants, sequences of multiple vowels and consonants (i.e. clusters, indicated in the Results sections by VOW and CON, respectively), syllable-based intensity peak intervals, and entire voiced (vcd) and voiceless (vcl) intervals. In addition, we included the total number (counts) of pauses, segments, and syllables (per presentation or speaker) as well as the speaking rate in syll/s in our rhythm analysis.

3. Results

3.1. Rhythm metric in charismatic vs neutral performances

We used a MANCOVA to investigate if and to what extent rhythmic parameters differed between the neutral and charismatic performances of the investor pitches. All traditional acoustic rhythm measures served as dependent variables. Presentation Style (neutral vs. charismatic) was the fixed factor. In addition, the speakers' Age, TOEFL score (for speaking), and Sex were included in the analysis as covariates. Their results are presented separately in 3.3. In 3.1, we begin with summarizing in Table 1 the significant findings of the fixed factor Presentation Style, based on additional pairwise comparisons tests carried out as part of the MANCOVA.

Table 1: *Test statistics of rhythmic differences between neutral ('neutr') and charismatic ('char') speeches. Parameters with non-significant results are omitted (e.g., Cons dur and speaking rate); ∅ = mean values.*

Rhythm feature	∅ neutr	∅ char	t	p	d
Pause count	3.2	4.4	2.2	0.04	0.7
Segment count	176.9	181.2	2.3	0.04	0.7
Syllable count	70.8	73.3	2.4	0.03	0.9
Vowel dur (ms)	83.1	95.5	3.4	0.003	0.7
ΔVOW (ms)	44.1	49.4	3.2	0.005	0.5
rPVI Vow	5.2	5.6	2.3	0.03	0.4
ΔCON (ms)	37.1	41.0	2.8	0.01	0.6
rPVI CON	3.9	4.6	2.9	0.01	0.8
Int. peak interval (ms)	244.8	262.4	2.2	0.04	0.6
vcd-interval dur (ms)	93.1	102.0	3.4	0.004	0.6
Δvcd interval (ms)	57.9	63.4	3.3	0.004	0.5
rPVI vcd interval	6.6	7.2	3.0	0.01	0.4

Table 1 shows that the speakers inserted more silent pauses into their charismatic performances and, thus, divided the presented pitch into smaller lexical chunks. At the same time, more sound segments and syllables were produced than in the neutral performance, meaning that the degree of speech reduction was lower (i.e. the presenters spoke more clearly) in their charismatic performances. Altogether, this points to a more listener-oriented speaking style in charismatic than in neutral pitches.

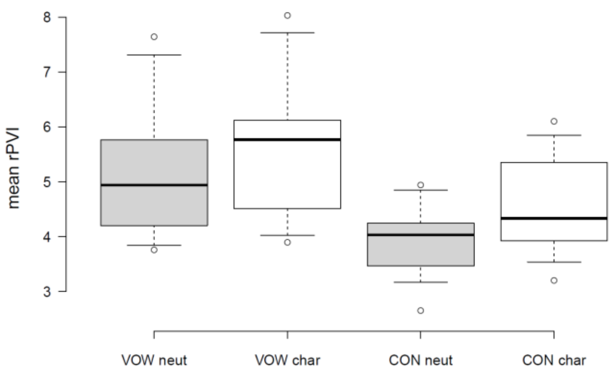

Figure 1: *rPVI results for vocalic (VOW) and consonantal (CON) speech intervals. The neutral and charismatic conditions are referred to as 'neut' and 'char', respectively.*

With regard to the rhythmic structuring itself, the differences between the two conditions of Presentation Style manifested themselves most clearly and consistently in the durations of individual vowels or entire voiced (vcd) intervals. In the charismatic performances these units became longer and more variable, see Table 1 and the rPVI example in Figure 1. That is, there were more extremely short and still more extremely long vocalic and voiced intervals in the charismatic as compared to the neutral presentation performances.

It is noteworthy that, at the same time, only the time intervals between intensity peaks increased in charismatic as compared to neutral speeches. The speaking rate itself remained the same for both Presentation Style conditions (which is why we used the rPVI instead of the tempo-normalized nPVI measures). That the speaking rate did not change is due to the consonantal and overall voiceless (vcl) intervals. On the one hand, they also be-came more variable in terms of increased Δ and rPVI measure-ments, see Table 1 and Figure 1. But, unlike the vowels and vcd intervals, the consonantal counterparts did *not* become overall longer in the charismatic than in the neutral performances. Rather, they became in tendency shorter, which means that the increase in the consonants' durational variability simultaneous-ly included a bias towards consonant reduction.

3.2. Correlations between rhythm measures and charisma

In 3.1., we determined how speakers changed the rhythmic characteristics of their speeches when being asked to switch from a neutral to a charismatic presentation mode (or vice versa). We found a dozen significant acoustic changes from neutral to charismatic. However, this does not automatically entail that these differences in speech rhythm are also picked up by listeners and, even if so, can then trigger systematic interpretations in terms of ratings of perceived speaker charisma. To address that question, we additionally calculated a series of correlations, based on the conservative Spearman's Rho correlation coefficient.

We took the difference between the average neutral and charismatic rating per speaker and correlated these values with the average rhythmic change from neutral to charismatic per parameter. That is, the correlation analysis was based on difference values, not on absolute values: $Rating_{char} – Rating_{neut}$ was correlated with $Measurement_{char} – Measurement_{neut}$ per speaker and parameter. Since we knew from the results in 3.1. in which direction the speakers changed their rhythm parameters from neutral to charismatic, were used one-tailed tests for the correlations.

Table 2: *Summary of significant correlations between rhythm measures and perceived speaker charisma.*

Rhyhtm feature	Spearmans ρ	p
ΔCON	0.41	0.050
rPVI CON	0.42	0.049
Δvcd interval	0.43	0.048
rPVI vcd interval	0.42	0.049

Table 2 shows that four correlations came out significant. All four correlations concerned measures of rhythmic variability, and all correlations were positive, see the example in Figure 2. That is, in general, what we found is that an increase in perceived speaker charisma is acoustically linked to an increase in rhythmic variability; or, in other words, what our results obviously suggests is that rhythmic variability is (particularly) beneficial for perceived speaker charisma. Note in this context that the correlations were slightly stronger for the duration variation among consonants or consonant sequences than for the duration variation among voiced speech sections, and that the vocalic variation itself was not correlated with charisma ratings. Both matches with the result that effect sizes in Table 1 were larger for consonantal than for vocalic or voiced rhythm units, see also the rPVI differences in Figure 1.

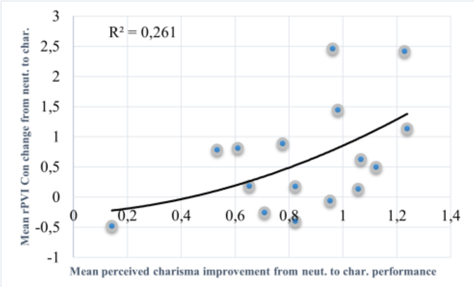

Figure 2: *Example correlation of rPVI CON with charisma.*

3.3. Effects of covariates

Due to space limitations, we cannot report all covariate test statistics here. In summary, however, the following results pattern emerged from the MANCOVA: The speakers' TOEFL score had a significant influence on the acoustic rhythm measures. A higher TOEFL score, i.e. a better L2 English competence, positively supported the rhythm changes from neutral to charismatic performance shown in Table 1. The differences between the two Presentation Style conditions became greater. This was particularly true for the mean (∅), ∆ and rPVI measures of VOW and vcd (all $p<0.01$), and only to a lesser extent also for ∆CON and rPVI CON ($p<0.05$).

A different result emerged for the covariate Age. Age effects were also significant, but associated with the speakers themselves. Thus, they applied in the same way to neutral and charismatic presentations. We found that older speakers generally had less clear pronunciation (in the form of a lower segment count, $p<0.01$) and a slower speaking rate (i.e. a lower syllable count, $p<0.001$, and higher mean values for VOW, CON and vcd, $p<0.05$). Evidence of less durational variability with increasing speaker age only manifested itself as a significant trend and, was moreover, limited to the vocalic measures ∆VOW ($p=0.08$) and rPVI VOW ($p=0.06$).

There were no influences on the magnitude of the neutral-to-charismatic rhythm differences for the covariate Sex. Like for Age, we only found general differences between the rhythmic (duration) structures of male and female speakers. Men showed an overall significantly higher speaking rate than women ($p=0.02$), combined with a less clear pronunciation (in the form of a lower segment count $p=0.04$) and shorter VOW and CON intervals ($p=0.01$, $p=0.003$). The rhythmic variability was, in tendency, also overall lower for men than for women, in particular in CON and VOW units (all p-values < 0.1).

4. Discussion and Conclusions

The research questions of our combined speech production and perception study were whether, firstly, systematic changes in traditional duration-based rhythm measures can be found when speakers change the style of their presentation performance from emotionally neutral and matter-of-fact to charismatic, i.e., an expressive, committed onstage presentation style. Secondly, we wanted to know if and in what way such potential rhythmic (i.e. sequential duration) changes also correlate with the degree of perceived speaker charisma. The speaker sample was moderately sized, but, because we used trained public speakers, we expect a high internal validity of our results.

The answer to the first question is a positive one. Results revealed that charismatic presentations exhibit amplified duration variations in consonantal and vocal intervals, alongside increased sonority and clearer speech. All three established types of rhythm measures contribute to the acoustic difference between the two presentation styles: average interval durations (Ø), standard deviations (Δ), and (raw) pairwise variability indices (rPVI). Along with the counts, 12 significant rhythmic differences emerged between the two presentation styles.

The answer to the second question is also a positive one. There are significant links between traditional duration-based rhythm measures and the degree of perceived speaker charisma. Duration-related aspects of speech rhythm are thus a relevant charisma factor, in line with indirect indications from previous empirical studies [24,25]. However, the links between production and perception only concern 4 of the 12 rhythm differences found; all of them related to either ∆ or rPVI measures. So what seems to be most important for perceived charisma is an increase in the speech rhythm's variability or contrastiveness. This aligns well with the conclusion of [25] that "more pronounced amplitude modulations biased raters toward higher perceived charisma ratings." (p. 176) Also in line with previous studies ([24,25]) is our finding that segment-size rhythmic elements play hardly any a role in the production and none in perception of a charismatic rhythm. Charisma-relevant rhythmic units rather seem to be cluster- and syllable-size – or larger.

In addition, while purely vocalic rhythmic elements mattered in production, they no longer played a role in the perception of charismatic presentation performances. Conversely, consonantal intervals, especially duration variations within consonant sequences (CON), gained greater importance in perception. This emerging asymmetry between production and perception needs to be further investigated and tested for its generalizability. This is of great practical relevance as it may mean that speakers, when trying to speak more charismatically, do not necessarily change their speech rhythm in a way that is most advantageous for them; and if this is true of trained public speakers, it is probably all the more true of naive speakers; [24] found that even Steve Jobs' speech rhythm differed from a baseline speaker sample especially in vocalic elements rather than in the consonantal elements that obviously matter more in charisma perception Thus, there is much to suggest that rhythm does not currently have the importance it should have in public-speaking training. Speakers need education on rhythm concepts and effective consonant- and variation-focused exercises to improve their delivery.

Regarding the tested covariates, the result with the greatest practical and conceptual relevance is that speaker sex can be neglected as a factor in the above considerations and conclusions. Differences between women and men are, if at all, quantitative but not qualitative. What must be considered, though, is the degree of language proficiency. Less fluent non-native speakers probably need more charisma-oriented rhythm training. This result also aligns with the relevance of "jaw dancing" exercises emphasized by [36] for both L2 learning and public speaking. Our follow-up steps consist of two activities: We will test which learning and feedback concepts are effective in training charisma-relevant rhythm factors and, moreover, we will repeat our study with untrained speakers.

5. References

[1] P. van Rijn, D. Poeppel, and P. Larrouy-Maestri, “Contribution of Pitch Measures over Time to Emotion Classification Accuracy,” *Preprint*, https://doi.org/10.31234/osf.io/pnysd, 2023.

[2] S. A. Almaghrabi, S. R. Clark, and M. Baumert, “Bio-acoustic features of depression: A review,” *Biomed. Signal Process. Control*, vol. 85, 105020, 2023.

[3] P. Chyan, A. Achmad, I. Nurtanio, and I. S. Areni, “A deep learning approach for stress detection through speech with audio feature analysis,” in *Proc. 6th International Conference on Information Technology, Information Systems and Electrical Engineering*, Yogyakarta, Indonesia, Dec. 2022, pp. 1-5.

[4] Q. Chen, W. Wang, Q. Zhang, and K. Yu, “Pre-training for spoken language understanding with joint textual and phonetic representation learning,” in *Proc. 22nd Interspeech Conference, Brno, Czech Republic. Sep.* 2021, pp. 1130-1134.

[5] M. Nagano, Y. Ijima, and S. Hiroya, “Perceived emotional states mediate willingness to buy from advertising speech,” *Frontiers in Psychology*, vol. 13, 1014921, Jan. 2023.

[6] Y. Amari, S. Okada, M. Matsumoto, K. Sadamitsu & A. Nakamoto, “Multimodal Analysis of Client Persuasion in Consulting Interactions: Toward Understanding Successful Consulting,” in G. Meiselwitz (ed.), *Social Computing and Social Media: Applications in Marketing, Learning, and Health.* Singapore: Springer, pp. 29-40, 2021.

[7] K. Fischer, O. Niebuhr, R. M. Langedijk, and S. Eisenberger, "Shall know you by your voice - Melodic and physical dominance in the design of robot voices," in *Proc. 1st International Seminar on the Foundations of Speech*, Sonderborg, Denmark, Sep. 2019, pp. 88-90.

[8] O. Niebuhr and J. Michalsky, "PASCAL and DPA: A pilot study on using prosodic competence scores to predict communicative skills for team working and public speaking," in *Proc. 20th Interspeech Conference*, Graz, Austria, Sep. 2019, pp. 306-310.

[9] O. Niebuhr, ”Advancing higher education practice by analyzing and training students vocal charisma: Evidence from a Danish field study,” in *Proc. 7th International Conference on Higher Education Advances*, Valencia, Spain, Jun 2021, pp. 1-8.

[10] J. Niehof, J. and O. Niebuhr, “What’s pitch got to do with it? – The connection between speech prosody and investor-pitch success,” in *Proc. 1st International Conference on Tone and Intonation (TAI)*, Sonderborg, Denmark, Dec 2021, pp. 288-292.

[11] P. Touati, “Prosodic aspects of political rhetoric,” in *Proc. of ESCA Workshop on Prosody*, Lund, Sweden, 1993, pp. 287-290.

[12] D. Duez, “Acoustic markers of political power,” *Journal of Psycholinguistic Research*, *26*, pp. 641-654. 1997.

[13] E. Strangert, and J. Gustafson, “What makes a good speaker? Subject ratings, acoustic measurements and perceptual evaluations” in *Proc. 9th Internspeech Conference*, Brisbane, Australia, Sep 2008, pp. 1688-1691.

[14] A. Rosenberg and J. Hirschberg, “Charisma perception from text and speech,” *Speech Communication*, vol. 51, no. 7, pp. 640-655, 2009.

[15] S. Scherer, G. Layher, J. Kane, H. Neumann, and N. Campbell, “An audiovisual political speech analysis incorporating eye-tracking and perception data,” in *Proc. 8th International Conference on Language Resources and Evaluation*, Istanbul, Turkey, 2012, pp. 1114-1120.

[16] O. Niebuhr, 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 *Proc. 9th International Conference on Speech Prosody*, Brisbane, Australia, Sep. 2018, pp. 359-363.

[17] O. Niebuhr and R. Skarnitzl,”Those who shout the loudest - Do they sound more charismatic?” in *Proc. 1st International Conference on Tone and Intonation (TAI)*, Sonderborg, Denmark, Dec 2021, pp. 1-5.

[18] O. Niebuhr and R. Skarnitzl, “Measuring a speaker's acoustic correlates of pitch - but which? A contrastive analysis based on perceived speaker charisma,” in *Proc. 19th International Congress of Phonetic Sciences*, Melbourne, Australia, Aug. 2019, pp. 1774-1778.

[19] K. J. Kohler, “Rhythm in speech and language: A new research paradigm,” *Phonetica*, vol. 66, no. 1-2, pp. 29-45, 2009.

[20] D. Appel Graber, *Verbal Behavior and Politics*. Urbana, IL: University of Illinois Press, 1976.

[21] D. Lakhani, *Persuasion: The art of getting what you want*. Hoboken: John Wiley & Sons, 2011.

[22] J. Apps, *Voice and Speaking skills for Dummies*. Hoboken: John Wiley & Sons, 2012.

[23] C. Bernhardt, *Nonverbal Communication in Recruiting*. Wiesbaden: Springer, 2022.

[24] O. Niebuhr, A. Brem, and J. Voße, “What makes a charismatic speaker? A computer-based acoustic-prosodic analysis of Steve Jobs' tone of voice,” *Computers in Human Behavior*, vol. 64, pp. 366-382, 2016.

[25] H. R. Bosker, “The contribution of amplitude modulations in speech to perceived charisma,” in B. Weiss, J. Trouvain, M. Barkat-Defradas, and J.J. Ohala (eds), *Voice Attractiveness: Studies on Sexy, Likable, and Charismatic Speakers*. Singapore: Springer Nature, 2021, pp. 165-181.

[26] A. Arvaniti, “Measuring speech rhythm,” in J. Setter & R.A. Knight (eds), *Cambridge Handbook of Phonetics*. Cambridge: Cambridge University Press, 2022, pp. 312-335.

[27] B. Bridgeman, D. Powers, E. Stone, and P. Mollaun, “TOEFL iBT speaking test scores as indicators of oral communicative language proficiency,” *Language Testing*, vol. 29, no. 1, pp. 91-108, 2012.

[28] J. Antonakis, N. Bastardoz, and P. Jacquart, "Charisma: an ill-defined and ill-measured gift," *Ann. Rev. Organ. Psychol. Organ. Behav.*, vol. 3, pp. 293-319, 2016.

[29] M. Simpson, “How to write a killer elevator pitch,” [Online]. Available: [Accessed: 15-Sept-2023]. URL: https://theinterviewguys.com/write-elevator-pitch/

[30] T. Kisler, U. Reichel, and F. Schiel, “Multilingual processing of speech via web services, ” *Computer Speech & Language*, vol. 45, pp. 326-347, 2017.

[31] P. Machač, and R. Skarnitzl, *Principles of phonetic segmentation*. Prague: Epocha, 2009.

[32] N. Taghva, A. Moloodi, V. Abolhasanizadeh, and R. Tabei, "A corpus study of durational rhythmic measures in the Kalhori variety of Kurdish," *Loquens*, vol. 10, no. 1-2, p. e098, 2023.

[33] F. Ramus, M. Nespor, and J. Mehler, "Correlates of linguistic rhythm in the speech signal," *Cognition*, vol. 73, no. 3, pp. 265-292, 1999.

[34] E. Grabe and E. L. Low, "Durational variability in speech and the rhythm class hypothesis," *Papers in Laboratory Phonology*, vol. 7, pp. 515-546. 2002.

[35] V. Dellwo, A. Leemann, and M. J. Kolly, “Rhythmic variability between speakers: Articulatory, prosodic, and linguistic factors,” *The Journal of the Acoustical Society of America*, vol. 137, no. 3, pp. 1513-1528, 2015.

[36] D. Erickson and O. Niebuhr, “Articulation of prosody and rhythm: Some possible applications to language teaching,” in *Proc. 13th International Conference of Nordic Prosody*, Sonderborg, Denmark, Jun 2023, pp. 1-45.

[37] O. Niebuhr, “F_0-Based Rhythm Effects on the Perception of Local Syllable Prominence,” *Phonetica*, vol. 66, no. 1-2, pp. 95-112.