MEASURING A SPEAKER'S ACOUSTIC CORRELATES OF PITCH – BUT WHICH? A CONTRASTIVE ANALYSIS BASED ON PERCEIVED SPEAKER CHARISMA

Oliver Niebuhr[1] & Radek Skarnitzl[2]

[1] University of Southern Denmark, Mads Clausen Institute, Centre for Industrial Electronics, Sønderborg/DK
[2] Institute of Phonetics, Faculty of Arts, Charles University, Prague/CZ
oniebuhr@mci.sdu.dk; radek.skarnitzl@ff.cuni.cz

ABSTRACT

Intonation is a multitasker. It conveys a plethora of non-lexical meanings and functions, and mirrors a speaker's personality traits, one of them being charisma. In the age in which leaders communicate globally through digital mass media, speaker charisma has become an important field of research. The number of different f0 level and range measures that are used, not only in phonetics, to determine speaker charisma call for standardization. Which f0 measures correlate most strongly with perceived speaker charisma and should thus be used in future analyses? Our study addresses this question based on 51 speakers whose charisma was rated by 42 listeners. Results show that the f0 mean is the best parameter to measure perceived charisma in terms of pitch level, while the best pitch distribution measures are kurtosis and the 80-percentile f0 range. The results are discussed in terms of gender differences and issues of cross-linguistic generalization.

Keywords: charismatic speech, leadership, f0, pitch perception, intonation.

1. INTRODUCTION

The fundamental frequency (f0) or its physiological or perceptual correlates are perhaps the parameter class with the greatest functional load in all languages of the world. Firstly, this is because voice (i.e., periodic vocal fold vibration) is the primary supplier of acoustic energy, without which speech communication over greater distances would not be possible. At the cost of a smaller acoustic richness, probably only whistling is able to bridge even greater distances than the voice [17].

Secondly, the generally high functional load of f0 reflects that speech is above all a facilitator of social interaction. Speech organizes hierarchy-building, negotiation, and group-formation processes and conveys emotions, attitudes, and discourse-control signals. These socio-communicative aspects of speech already existed at a time when recursive syntax and a sound-segment-based lexicon were far beyond the hominid's anatomical and cognitive abilities [7,9,26]. Until today, these socio-communicative aspects are still mainly shaped by speech melody [1,8,27,29]. Also, listeners give melodic meanings priority if they are in conflict with lexical ones [12].

Thirdly, the functional load of f0 is due to the fact that f0 is additionally and consistently a mirror of physiological processes and, thus, directly linked to certain socio-communicative indicators of speaker state. These include the emotional, health and mental conditions of a speaker, including pain, stress and physical exercise [13,14,16].

Accordingly, speakers of a language may, for instance, choose not to use [f] or [ø] functionally (i.e., distinctively); or they may choose not to use final or segmental lengthening functionally; but they cannot decide against using f0 functionally. This fact is also reflected in phonetic research: 78.3 % of the papers at the last International Conference of Speech Prosody (2018) directly or indirectly dealt with f0 and its phonetic correlates. At the last International Congress of Phonetic Sciences (2015), 17.6 % of the papers addressed f0, although the scope of the congress goes far beyond prosody alone.

Taking into account the central relevance of f0 in communication, phonetic research has developed and applied a plethora of acoustic measures to determine the melodic characteristics of speakers. The *pitch level* characteristics of a speaker can be measured, among other things, using absolute measurements such as the mean or median f0, or relative measurements in which the absolute ones are related to a reference or correction value. A common reference is the overall span of f0 values used by a speaker; a frequent correction value is the 7.64th percentile, by which the mean or median f0 becomes more representative of the speaker's f0 baseline [15].

The *pitch range* characteristics of a speaker can also be determined in various ways. These include the f0 range (the difference between absolute minimum and maximum), the f0 standard deviation, the coefficient of variation (varco), and the 80-percentile range that corresponds to the difference between the 90th and 10th percentile. In the forensic phonetic domain, it has been suggested that the distribution of f0 values is also reflective of the given speaker [10]; f0 distributions may be parameterized by, for instance, spectral moments [25].

Given all these f0 measures, which ones are best? There is not a single answer to that question. The best measures in scientific terms are those that are robust (i.e. context independent) and reliable (i.e. detectable and not error-prone). The present paper takes an application-oriented perspective, specifically that of a prosody-based rhetorical training. From this perspective, measures additionally have to be explainable, user-friendly and closely correlated with perceived speaker traits.

The trait we focus on in this study is perceived *speaker charisma*. Numerous previous production and perception studies showed that pitch level and pitch range both play a crucial role in perceived speaker charisma [8,22,24], the range probably even more so than the level [3]. Moreover, these studies showed that pitch level and range can be trained and improved and, thereby, also a speaker's charismatic impact on listeners [19,2].

However, what is currently still lacking is a detailed understanding of how well the individual acoustic measures of pitch level and range correlate with perceived speaker charisma. This understanding is essential if, for example, automatic assessment and training algorithms are to be developed which quantify the charismatic impact of a speaker's speech signal and, on this basis, provide feedback to the speaker as to how s/he can improve this impact. Such feedback could be given in the form of real-time visualizations of a speaker's pitch-level and/or pitch-range information [20], which is either color-coded (red=bad, green=good) or displayed in relation to certain target or reference values.

A pilot study [18] has already found indications that the f0 median is more strongly correlated with perceived speaker charisma than the f0 baseline (7.64^{th} percentile), and that the f0 percentile range could be more closely related with perceived speaker charisma than the f0 standard deviation. However, the speaker sample in that study was small, and only a limited set of basic pitch-level and pitch-range measures were tested. Therefore, the present study extends this pilot correlation analysis by including more than four times as many speakers as data points in the correlations and by expanding the set of tested acoustic measures beyond a few basic pitch level and pitch range parameters.

Five research questions are addressed: (1) Are there differences in how well the various level, range, and distribution measures of pitch correlate with perceived speaker charisma? (2) Which pitch-level measure correlates most strongly with perceived speaker charisma? (3) Are parameters pertaining to the shape of f0 distribution related to perceived speaker charisma? (4) Which pitch-range measure correlates most strongly with perceived speaker charisma? (5) Do the answers to (1)-(4) depend on speaker sex?

2. METHOD

2.1. Material

The analysis is based on 51 post-graduate students (33 males, 18 females), aged 23–27 years, from the Dept. of Technology Entrepreneurship and Innovation at the University of Southern Denmark (www.sdu.dk/tei). All were highly proficient non-native speakers of English at B2 or C1 levels (according to university internal entry tests).

For the purpose of our recordings, the speakers held a so-called "elevator speech", i.e. a short description of a start-up company or its underlying business idea presented with the intention to acquire external funding from investors. The 51 speakers were trained and gained experience in giving these elevator speeches during their business engineering or entrepreneurship educations. All recordings were made while the speakers presented charismatically and persuasively as they could in front of a real audience. The recordings took place in the sound-treated SDU MCI Innovation Lab at the University of Southern Denmark in Sønderborg. The speakers were standing while speaking, see Figure 1.

Figure 1: Example of an elevator speech.

We used 40–98 s presentation excerpts (ø 61.4 s) for our analyses, cut out at major prosodic-phrase boundaries from the middle of each elevator speech.

2.2. Acoustic analysis

The acoustic analysis was performed in Praat [5], using autocorrelation [4] with a 40 ms window length; the f0 of male voices was extracted in a frequency range of 75–320 Hz, that of female voices in a frequency range of 75–480 Hz. Per speaker/ presentation excerpt, we extracted three measures of pitch level and four measures of pitch range:

- Pitch-level: *mean, median, baseline*;
- Pitch-range: *standard deviation, varco, range, 80-percentile range*.

In addition, f0 was calculated every 5 ms across each speaker's presentation excerpt and two f0 distribution measures were derived from each resulting set of values: *skewness* (symmetry around the mean) and *kurtosis* (sometimes described as peakedness, but more appropriately reflecting the thinness or fatness of the distribution tails). The *moments* package [11] in R [23] was used for this analysis. In total, we determined nine f0-based pitch measures per speaker/presentation excerpt.

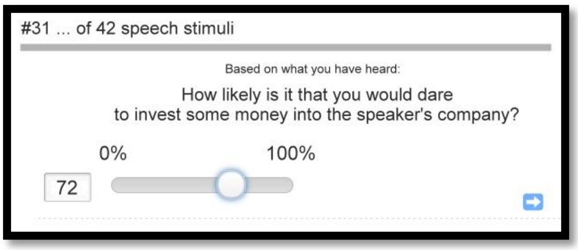
2.4. Listener ratings of perceived speaker charisma

The mean values of each acoustic measure were correlated with the perceived charisma levels of the 51 speakers using Pearson's product moment correlation coefficient. Perceived speaker charisma was rated based on the question "On a scale from 0-100%, how likely is it that you would dare to invest some money into the speaker's company?" A pilot study found this question to be easily applicable by listeners [18]. Among all questions asked in the pilot study, the one on investment likelihood turned out to be the one whose responses varied most strongly as a function of f0 measures.

The elevator-speech excerpts were delexicalized through low-pass filtering (in order to exclude additional noise in the ratings from lexical contributions to speaker charisma, [2]) and then integrated in an online perception experiment based on SurveyXact. The speech excerpts were presented to listeners in individually randomized orders. Ratings were made by shifting a slider on the screen to the respective percentage for each stimulus (Figure 2).

For the sake of simplicity in this short paper, we will directly translate investment likelihood ratings into perceived speaker-charisma ratings. We do this because previous studies showed that investment likelihood is closely associated with key attributes of charisma, such as being decisive, performing, inspiring, and persuasive [24,28].

Figure 2: Screenshot of online rating task.



A total of 42 listeners participated in the rating task. All listeners were proficient non-native speakers of English, like the speakers they listened to.

3. RESULTS AND DISCUSSION

Before we discuss the correlations between speaker charisma ratings and the individual f0 measures, two other important aspects of our results must be noted. First, the mean charisma scores did not differ between male and female speakers (t-test: $t(49) = 0.28$; p=0.78). Second, although the duration of the speech samples was relatively variable (see section 2.1), this had no effect on the ratings: (r = –0.018, p = 0.90).

3.1. Central tendency of f0

Correlations between central tendency measures and listener ratings are shown in Table 1. The arithmetic mean of f0 values appears to be most suitable for capturing the respondents' judgments of speaker charisma, with all correlations being significant at p<0.05. Interestingly, the f0 baseline, while being most effective for capturing a speaker's individuality across different speaking situations [15], is not well correlated with ratings of speaker charisma.

Table 1: Correlations (r) of f0 central tendency measures with speaker-charisma ratings, N=51. ** p <0.05; * p <0.1

	all	female	male
mean	0.35 **	0.50 **	0.61 **
median	0.36 **	0.45 *	0.60 **
baseline	0.20	0.21	0.36 **

All the correlation scores shown in Table 1 are positive; this means that, for both female and male speakers, higher f0 means seem to be associated with higher charisma ratings (see the discussion section for further comments on gender aspects).

3.2. Variability of f0

Correlations of variability measures with charisma ratings are shown in Table 2. The superiority of the 80-percentile range is obvious: correlation with charisma ratings is strong and highly significant for all speakers, as well as for males and females separately. In contrast, the traditional f0-range measure, i.e. the difference between maximum and minimum f0, shows no correlation with the ratings. This is probably because this traditional range measure is not very robust against typical octave errors in f0 extraction: it is sufficient that one very low/high f0 value is (erroneously) identified for the range to be affected dramatically. Figure 3 shows scatterplots of the two measures, as compared with charisma ratings.

Note that the coefficient of variation (varco; i.e., the standard deviation related to the mean) does not correlate well with charisma ratings, only in female speakers. The reason for this is not readily apparent, but it would be worth being investigated further.

Regarding the f0-distribution measures, skewness and speaker charisma are negatively correlated. This makes perfect sense in the light of previous findings [21,22]. The better a speaker is able to shift the majority of his/her f0 values to the top of the individual f0 range, or, in other words, the more often and longer a speaker's voice shoots upwards from the bottom of the f0 range, the more beneficial this is for his/her perceived charisma. This correlation applies to all speakers, but less so for females, perhaps because their mean f0 already lies relatively higher within the f0 extraction range than that of males.

Table 2: Correlation (*r*) of f0 variability and distribution measures with speaker-charisma ratings, N=51. *** p <0.001; ** p <0.05; * p <0.1

	all	female	male
SD	0.35 **	0.77 ***	0.26
varco	0.02	0.47 **	-0.09
range	0.04	0.10	0.02
80-perc. range	0.52 ***	0.71 ***	0.62 ***
skewness	–0.42 **	–0.13	–0.53 **
kurtosis	–0.43 **	–0.70 **	–0.49 **

Figure 3: Correlation of charisma ratings with the 80-percentile f0 range (left) and f0 range (right).

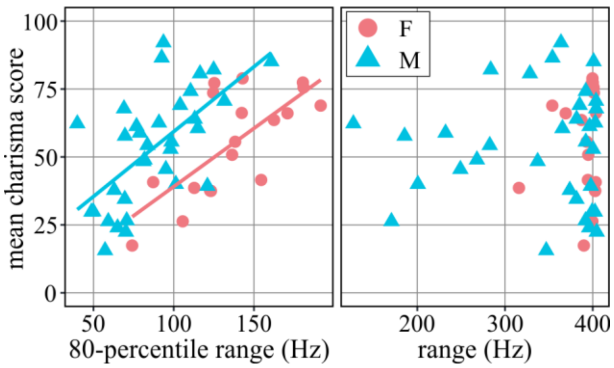
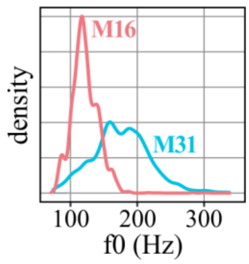


Figure 4: F0 distribution of two speakers illustrating kurtosis differences (see text).



Kurtosis is also negatively correlated with charisma ratings and is, unlike skewness, a more gender-robust f0-distribution measure. The more a speaker is able to make equal use of his/her entire f0 range, the better this is for perceived charisma. Figure 4 compares the f0 distribution for speaker M16 with the highest kurtosis value (investment likelihood 29.8%) and speaker M31 with kurtosis approaching the mean for male speakers (investment likelihood 74.2%). Accordingly, there is a strong negative correlation between kurtosis and the 80-percentile range: $r = –0.53$, $p < 0.001$.

4. CONCLUSIONS

Our study demonstrated on the example of perceived speaker charisma that the choice of f0 measures does make a big difference. Mean f0 and baseline f0 are seemingly equivalent and even interchangeable measures; the same applies to f0 range and f0 standard deviation. But, that is clearly not true. Especially when it comes to measuring meaningful attitudinal or expressive speaker traits or similar paralinguistic aspects of communication, a proper selection of f0 measures is of great importance. We must particularly point to the danger of using the traditional f0 range as an indicator of speaker traits like charisma: our results show no correlation between listener ratings and the f0 range. Conversely, the 80-percentile range, which is cleared of extreme values due to errors in f0 extraction, captures the listeners' ratings of speaker charisma best.

Based on the present data, our recommendation is to use the arithmetic mean for capturing a speaker's pitch level, and to use the 80-percentile range, perhaps along with kurtosis, for representing a speaker's pitch variability. The recommendation for these general pitch measures holds over and above languages and speaker charisma, whereas the findings on skewness, varco etc. might be more specific to speaker charisma. In general, our results also suggest that it is advantageous to analyze male and female speakers separately, not only because their f0 values lie in different ranges, but also because the correlations between speaker traits like charisma and some of the parameters vary considerably as a function of speaker sex. The explanations for this may be obvious for some parameters like skewness, but less so for others like, for instance, the f0 baseline.

Finally, our results differ in some respects from previous findings. For instance, previous studies showed that males sound more charismatically than females [6]. Our data, however, come from recordings of speakers who received intensive charismatic-presentation training; and it has been shown that this training is able to level out gender differences in the ears of listeners [19]. Studies also found a positive correlation between stimulus duration and charisma rating [24]. They used very short stimuli, though (<10s). We assume that this duration effect has an upper threshold (appr. 20s) and that our stimuli were too long for this effect to still emerge.

The time domain is also the direction, in which we will extend the present line of research. We proceed with testing how valid and robust the f0 measures are against changes in speech-sample duration. For example, compared to the 80-percentile range, kurtosis might need a longer stretch of speech (i.e. more f0 values) to yield proper measurements.

5. REFERENCES

[1] Abelin, A., Allwood, J., 2000. Cross linguistic interpretation of emotional prosody. *Proc. ISCA Workshop on Speech and Emotion 1*, 110–113.

[2] Antonakis, J., Fenley, M., and Liechti S. 2011. Can charisma be taught? Tests of two interventions. *Acad. Manag. Learn. Educ. 10*, 374–396.

[3] Berger S., Niebuhr O., Peters B. 2017. Winning Over an Audience – A Perception-based Analysis of Prosodic Features of Charismatic Speech. *Proc. 43rd Annual Conference of the German Acoustical Society, Kiel, Germany*, 1454–1457.

[4] Boersma, P. 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *IFA Proceedings* 17, 97–110.

[5] Boersma, P., Weenink, D. 2018. *Praat: doing phonetics by computer* (Version 6.0.43). Retrieved from http://www.praat.org on October 1, 2018.

[6] Brooks, A.W., Huang, L., Kearney, S.W., Murray, F.E. 2014. Investors prefer entrepreneurial ventures pitched by attractive men. *Proc. National Academy of Sciences of the United States of America (PNAS) 111*, 4427-4431.

[7] Brown, S. 2017. A Joint Prosodic Origin of Language and Music. *Frontiers in Psychology 8*, 1894. https://doi.org/10.3389/fpsyg.2017.01894.

[8] D'Errico F., Signorello R., Demolin D., Poggi I. 2013. The perception of charisma from voice. A crosscultural Study. *Proc. Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, Switzerland*, 552–557.

[9] Falk, D. 2004. Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences 27*, 491-503.

[10] Kinoshita, Y., Ishihara, S., Rose, P. 2009. Exploring the discriminatory potential of F0 distribution para-meters in traditional forensic speaker recognition. *Int. J. of Speech, Lang. and the Law 16*, 91–111.

[11] Komsta, L., Novomestky, F. 2015. *Moments* http://www.rdocumentation.org/packages/moments

[12] Landgraf, R. 2014. Are you serious? Irony and the perception of emphatic intensification. *Proc. 4th International Symposium on Tonal Aspects of Languages, Nijmegen, The Netherlands*, 91-94.

[13] Landgraf, R., Schmidt, G., Köhler-Kaeß, J., Niebuhr, O., John, T. 2017. More Noise, Less Talk - The Impact of Driving Noise and ICC Systems on Acoustic-prosodic Parameters in Dialogue. *Proc. 43rd DAGA conference, Kiel, Germany*.

[14] Lautenbacher, S., Salinas-Ranneberg M., Niebuhr, O. 2017. Phonetic characteristics of vocalizations during pain. *PAIN Reports 2*, e597.

[15] Lindh, J., Eriksson, A. 2007. Robustness of long time measures of fundamental frequency. *Proc. Interspeech 2007* Antwerpen, 2025–2028.

[16] Marquard, C., Baasch, C., Brodersen, M., Niebuhr, O., Schmidt, G. 2017. Speech, Think, Act: A Phonetic Analysis of the Combinatorial Effects of Respiratory Mask, Physical and Cognitive Stress on Phonation and Articulation. *Proc. 43rd DAGA conference, Kiel, Germany*, 1-5.

[17] Meyer, J. 2008. Typology and acoustic strategies of whistled languages: phonetic comparison and perceptual cues of whistled vowels. *J. Int. Phon. Assoc. 38*, 69-94.

[18] Niebuhr, O., Skarnitzl, R., Tylečková, L. 2018. The acoustic fingerprint of a charismatic voice - Initial evidence from correlations between long-term spectral features and listener ratings. *Proc. 18th International Conference of Speech Prosody, Poznán, Poland*, 359-363.

[19] Niebuhr, O., Tegtmeier, S., Schweissfurth, T. under review. Female speakers benefit more than male speakers from prosodic charisma training – A before-after analysis of 12-week and 4-hour courses. *Frontiers in Communication*.

[20] Niebuhr, O., Tegtmeier, S., Brem, A. 2017. Advancing research and practice in entrepreneurship through speech analysis – from descriptive rhetorical terms to phonetically informed acoustic charisma metrics. *J. of Speech Sciences 6*, 3-26.

[21] Niebuhr, O., Thumm, J., Michalsky, J. 2018. Shapes and timing in charismatic speech – Evidence from sounds and melodies. *Proc. 9th International Conference of Speech Prosod, Poznań, Poland.*

[22] Niebuhr, O., Voße, J., and Brem, A. 2016. What makes a charismatic speaker? A computer-based acoustic prosodic analysis of Steve Jobs tone of voice. *Computers and Hum. Beh. 64*, 366–382.

[23] R Core Team. 2016. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria http://www.R-project.org

[24] Rosenberg, A., Hirschberg, J. 2009. Charisma perception from text and speech. *Speech Communication 51*, 640–655.

[25] Skarnitzl, R., Hývlová, D. 2014. Statistický popis hodnot základní frekvence. In: Skarnitzl, R. (ed), *Fonetická identifikace mluvčího*. Praha: FF UK, 49–64.

[26] Tomasello, M. 2008. *Origins of Human Communication*. The MIT Press. Cambridge, MA.

[27] Vaissière, J. 1995. Phonetic explanations for cross-linguistic similarities. *Phonetica 52*, 123-130.

[28] Weninger, F., Krajewski, J., Batliner, A., Schuller, B., 2012. The Voice of Leadership: Models and Performances of Automatic Analysis in On-Line Speeches. *IEEE Transactions on Affective Computing 3*, 496-508.

[29] Yanushevskaya, I., Gobl, C., Ní Chasaide, A. 2018. Cross-language differences in how voice quality and f0 contours map to affect. *JASA 144*, 2730-2750.

[1] The second author was supported from the European Regional Development Fund-Project "Creativity and Adaptability as Conditions of the Success of Europe in an Interrelated World" (No. CZ.02.1.01/0.0/0.0/16_019/0000734).