Digital Rhetoric 2.0: How to Train Charismatic Speaking with Speech-Melody Visualization Software

Niebuhr, Oliver; Neitsch, Jana

Published in:
Speech and Computer - 22nd International Conference, SPECOM 2020, Proceedings

DOI:
10.1007/978-3-030-60276-5_35

Publication date:
2020

Document version:
Accepted manuscript

Citation for pulished version (APA):
Niebuhr, O., & Neitsch, J. (2020). Digital Rhetoric 2.0: How to Train Charismatic Speaking with Speech-Melody Visualization Software. In A. Karpov, & R. Potapova (Eds.), *Speech and Computer - 22nd International Conference, SPECOM 2020, Proceedings: 22nd International Conference, SPECOM 2020, St. Petersburg, Russia, October 7–9, 2020, Proceedings* (pp. 357-368). Springer. https://doi.org/10.1007/978-3-030-60276-5_35

Digital rhetoric 2.0: How to train charismatic speaking with speech-melody visualization software

Oliver Niebuhr [0000-0002-8623-1680] and Jana Neitsch [0000-0002-2185-8829]
University of Southern Denmark, MCI/CIE, Alsion 2, 6400 Sønderborg, Denmark
olni@sdu.dk

Abstract. The present study deals with a key factor of speaker charisma: prosody or, to put it less technically, speech melody. In a contrastive analysis, we investigate the extent to which computer-aided real-time visualizations of acoustic-melodic parameters in a self-guided training task help speakers use these parameters more charismatically. Fifty-two speakers took part in the experiment, subdivided into four equally large, gender-balanced groups. Each group received the same introduction and instruction, but then trained their melodic charisma with a different software tool. One tool – the "Pitcher" – was specifically developed for visualizing melodic parameters and providing feedback on their proper use in charismatic speech. The other software tools serve language- or singer-training purposes. Results show that three out of the four tools in fact support speakers in adopting a more charismatic speech melody after training, whereby the "Pitcher" outperforms all other tools by more than 25%. We discuss the implications of our findings for refining speech-melody visualization strategies (also for L2 training) and for digitizing rhetorical speaker coaching.

Keywords: Charisma, Pitcher, voice, prosody, public speaking, rhetoric.

1 Introduction

1.1 What is charisma?

Everyone has already intuitively experienced charisma in everyday situations. At the same time, however, charisma is such a complex and versatile phenomenon that those who experience it can hardly explain it or assign it to a specific perceptual feature or personality trait. This is the reason why charisma has been described as "easy to sense but hard to define" [1:305] in previous studies.

Nevertheless, at least since the Aristotle era, researchers have tried to itemize, systematize and, ultimately, to objectify charisma. Aristotle himself already pointed out that a speaker’s persuasive impact on others critically depends on emotions, values and symbolic language. However, the term 'charisma' was not introduced by him, but by the German sociologist Max Weber in the early 20th century [2]. Moving away from Aristotle's idea, Weber defined charisma as an extraordinary and almost magical gift that is innate in a speaker. By this, he turned charisma into a categorical concept. That is, Weber thought of charisma as a feature that a speaker either does or does not have. In contrast, modern definitions of charisma, derived from decades of empirical and experimental research in politics, psychology, and management, take a few steps back to Aristotle's original idea. Charisma is nowadays considered neither innate nor magical anymore. Rather, it is defined as a skill that relies on concrete, measurable signals and that, like any other gradually pronounced skill, can be learned and improved by its users and assessed by its trainers.

In terms of the influential charisma definition by Antonakis [3] and a recent terminological refinement by Michalsky and Niebuhr [4], charisma represents a particular communication style. It gives a speaker leader qualities through symbolic, emotional, and value-based signals. Three classes of charisma effects are to be distinguished in this context: (i) conveying emotional involvement and passion inspires listeners and stimulates their creativity; (ii) conveying self-confidence triggers and strengthens the listeners' intrinsic motivation; (iii) conveying competence creates confidence in the speakers' abilities and hence in the achievement of (shared) goals or visions. Inspiration, motivation, and trust together have a strongly persuasive impact by which charismatic speakers are able to influence their listeners' attitudes, opinions, and actions.

1.2 A charismatic speech melody and its impact on listeners

Since charisma has been defined as a way of communicating, studies investigated the relative contributions of verbal and nonverbal communication features on a speaker's charismatic impact. As it turned out, speech melody – i.e. the complex interplay of intonation, stress, rhythm, loudness and voice quality in the acoustic signal – is anything but a negligible factor for perceived speaker charisma [5,6]. As Amon puts it: "there is a clear superiority of the audible impression over the visible" [7: 19-20] (our translation).

A number of recent studies support Amon's statement and demonstrate the strong effects of a charismatic speech melody on listeners. For instance, [8] investigated the acoustic characteristics of the US presidential candidates between 1960 and 2000. To that end, they analyzed the famous American TV debates that always take place between the two remaining opponents of the republican and democratic parties just before the actual election. From the extracted signals that were low-pass filtered to cover the range from f0 to roughly the first formant, they calculated a measure of the spectral energy distribution. This measure was able to give a 100% accurate prediction of the outcomes of all eight US presidential elections over the forty-year period.

Similarly, [9] analyzed the speech-melody profiles of 175 founders who presented their business ideas to investors at major startup events. Results showed that about 80% of the investors' funding/no-funding decisions could be explained by speech-melody patterns alone. In an earlier study, [10] and [11] used such speech-melody patterns to evoke striking effects on people's thoughts and behavior. For instance, in experiments with talking robots, [10] and [11] superimposed the exceptionally charismatic speech melody of Steve Jobs or the moderately charismatic speech melody of Mark Zuckerberg on a robot's synthesized speech, all else equal, including the spoken words. Results showed that participants stuck to what the robot with Jobs' speech melody said and largely disregarded suggestions or commands made by the robot with Zuckerberg's speech melody. For example, Jobs' speech melody made participants fill out longer questionnaires or book different sightseeing trips; and implemented in a car-navigation system's voice, it made the participants follow a suggested route for longer, even if this route meant a considerable detour for them.

1.3 The Pitcher: A software for melodic charisma training

Regarding 1.2, we have now reached a point in science at which charismatic effects in terms of changes in listeners' thoughts and actions have become a learnable, improvable, and measurable skill. Moreover, we have arrived at a solid understanding of how speech melody, as a key element of charisma, contributes to these changes in thoughts and actions. This includes that we know where the "sweet spots" are located, i.e. narrow charisma-enhancing target-value ranges that speakers have to hit along each parameter of their speech melody [12]. In other words, we can start asking how we can train speakers such that they adopt a more charismatic speech melody.

In this context, it quickly became apparent that visualizing speech melody such that it becomes instructive and implementable for speakers is not a trivial task. Different existing visualization methods developed for research or second-language teaching have been modified, combined, and tested by [13]. Results showed that, on the one hand, speakers cannot cope with abstract, symbolic representations. On the other hand, graphical representations with many melodic or rhythmic details easily confuse speakers, provoke them to make mistakes, and slow down their learning [13]. The most instructive and effective representation of speech melody consists of an "angular" pitch contour that is stylized at major intonational landmarks (like pitch-accented syllables) and a stress pattern that is displayed separately and reduced to two or three stress levels, i.e. stressed/no stress or no stress/partial stress/full stress [14]. Note that we use the term 'stress' here in the sense of concrete perceptual prominence [15].

Fig. 1. Three screenshots of the Pitcher software with dots embedded in the pitch contour (top) as well as with pitch contour and dots displayed separately for focused learning (bottom).

Based on these findings, we developed the "Pitcher" – a tool for the computer-aided learning of a more charismatic speech melody (see Fig. 1). The Pitcher measures major charisma-related acoustic parameters of a speaker's speech melody, for example, while giving a presentation. Based on these measurements, the Pitcher provides simple, color-coded real-time feedback for the individual parameters, with "green" and "red" indicating whether the respective parameter is within or outside its charisma-enhancing "sweet-spot" target range. Target ranges are defined gender-specifically. The parameters taken into account by the Pitcher are: f0 range, mean f0 level, lowness of phrase-final falling f0, duration of prosodic phrases or interpausal units (IPUs), silent pause duration, speaking rate, and acoustic energy.

Furthermore, dots can be shown in the Pitcher display. Each dot marks a syllable. Depending on the display mode chosen by the speaker, the dots provide feedback on speaking rate or acoustic energy. In the speaking-rate mode, they change their color from green through yellow to red if the speaker becomes too fast or slow in his/her speech. In the acoustic-energy mode, the dots additionally change their size, with larger dots indicating a higher acoustic-energy level. If speakers get too soft in their voice, the smaller dots turn red. Note that speakers are able to hide all dots in order to focus on the f0-related feedback only. Inversely, they can hide the stylized f0 contour in order to focus on the dot-related tempo or loudness feedback only.

Finally, feedback on IPU and silent-pause durations is provided by means of the Pitcher window itself that represents the time axis. Both the pitch contour and the dots emerge over time from left to right along this time axis. The time axis is chosen such that the right end of the Pitcher window coincides with the end of the charisma-enhancing IPU duration. The time axis is reset when the speaker makes a silent pause larger than 500 ms. Thus, the parameter targets for IPU duration and silent pause duration are reached when speakers constantly insert pauses >500 ms in their speech before they have reached the right end of the Pitcher window.

1.4 Research questions and aims

With the present study, we aim to test within a controlled experimental setting whether the Pitcher works effectively in that it helps speakers learn to produce a more charismatic speech melody. Moreover, we aim at determining whether the Pitcher is in this respect more effective than other speech melody visualization tools that are not specifically designed to enhance speaker charisma. By addressing these points, we also aim to better understand at a more general level how software tools like the Pitcher manage to cause systematic changes in a speaker's speech melody at all and whether some acoustic parameters are more or less susceptible to this change.

Fig. 2. Screenshots of the three software tools to which the Pitcher is compared.

The other melody-visualization tools to which the Pitcher is compared are (1) "See & Sing Professional 1.5.8", (2) "AmPitch 1.2", and (3) "RTPitch 1.3a", see Fig. 2. Tool (1) displays f0 as a continuous non-stylized line that is, like for all others tools, drawn from left to right across the screen. It is automatically reset after 5 s and stops during silent pauses. F0 contours are associated with sequences of musical notes and, in addition, a separate vertical acoustic-energy meter is shown on the right edge of the screen. Tool (2) displays f0 by a discontinuous line (interrupted by silent pauses and voiceless consonants). The representation of acoustic energy is integrated in the thickness of that line. Unlike for tool (1), this line is not automatically reset for 60 seconds (and horizontal scrolling was switched off as well). Thus, tool (2) creates an overview of the frequency distribution of a speaker's f0. For tool (3), the horizontal-scrolling function is mandatory, but, like for tool (2), the f0 contour is drawn in a discontinuous, non-stylized fashion. Acoustic-energy information is provided indirectly by the waveform amplitudes that are displayed in a separate window at the top.

Based on the comparison of these four different software tools (i.e., the Pitcher + tools (1)-(3)), the following questions are addressed in the present paper:

- Can speakers learn to adopt a different speech melody through software-supported real-time visualizations of acoustic-melodic features?
- If so, does the Pitcher outperform the other visualization tools?
- Does the comparison between the four tools provide additional indications about
  - ... how the difference between a continuously and a discontinuously displayed pitch contour affects the learning of melodic parameter settings?
  - ... how the difference between pitch-contour reset and horizontal scrolling affects the learning of melodic parameter settings?
  - ... how the different ways of displaying acoustic energy affect the learning of melodic parameter settings, a strong and loud voice in particular?

2 Method

2.1 Participants

Overall, 52 participants took part in the study (19 female, 33 male, all between 19 and 22 years old). All of them were Bsc students of electrical engineering in their first semester at the Mads-Clausen Institute of the University of Southern Denmark (SDU) in Sønderborg. All participants were proficient L2 English speakers, i.e. at level B2 or higher on the European CEFR scale [16] according to SDU-internal entry tests. Furthermore, no participant had former experience in public speaking or in using any of the tested/compared software tools. There were also no (semi)professional singers among the students and none of the students reported any speech or hearing disorders.

2.2 Procedure

The experiment began in a lecture hall at SDU (Fig. 3). All 52 students received a 90-minute plenary lecture about speaker charisma and the specific key characteristics of a charismatic speech melody (see 2.3). The lecture, given by the first author, has been successfully tried and tested in numerous professional charisma coachings with managers, entrepreneurs, and sales agents by the consulting company AllGoodSpeakers ApS of the first author. After this introduction, participants presented, in a random order, a short elevator pitch in front of their fellow students. An elevator pitch is a short outline of a business idea meant to attract customers and/or investors. The given elevator pitches were about 2-3 minutes long and recorded with a head-mounted mic (Sennheiser HSP2-3) and a Zoom-H4 recorder (48 kHz, 24-bit). The 52 recorded pitches constituted the before-training condition of the experiment.

In the next step, the participants were randomly split up into 4 groups of 13 speakers each. Speaker gender was balanced across the groups such that there were 4-5 female speakers per group. Each group was instructed to practice the elevator pitch according to the melodic key characteristics that they had learned in the lecture, using – as an additional resource – a software tool that would visualize their speech melody in real-time. Based on this instruction, each group was assigned to *one* of the four software tools (Speech & Sing, AmPitch, RTPitch, Pitcher) and *not* allowed to use any other or additional tool for their training. Then, the group members were sent out in order to distribute themselves across the campus for individual self-guided training, having the respective software tool on their student laptop.

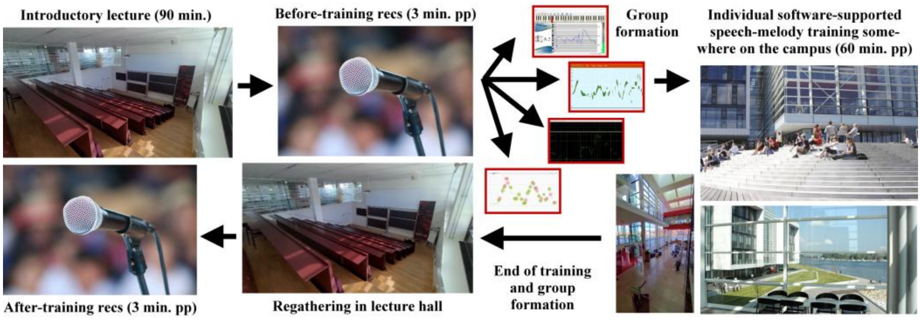

Fig. 3. Steps of the experimental procedure (photos by SDU, Oliver Niebuhr, pxfuel.com).

After one hour of this individual software-supported presentation training, all students met again in the lecture hall. In a differently randomized order, they held the same elevator pitches as before and were again recorded with the same equipment as before. These further 52 recordings constituted the experiment's after-training condition.

Note that in the chosen experiment design, the before/after-training effect is to a certain extent confounded with a familiarization effect that concerns the task and the presented text. The reason why we nonetheless deliberately chose this setup is that, firstly, this familiarization effect inevitably also occurs in real training scenarios. Our design is hence ecologically valid. Secondly, familiarization effects are not simply positive for speaker charisma. For one thing, familiarization means less stress and, thus, a lower f0 level [17], whereas a higher f0 level is required for charismatic speech [4]. For another thing, [18] have shown that intensive presentation training quickly causes routine/boredom, which also has unfavorable effects on the melodic aspects of speaker charisma. Therefore, it is reasonable to ask whether and which training software can improve melodic speaker charisma, also and in particular in the chosen experiment design. Another point is that the determination of the relative performance of the "Pitcher" software compared to the other three software tools is a between-subjects matter in the chosen experiment design. Thus, any potential within-subjects familiarization effects have no influence on this important question.

2.3 Data analysis

The 2 x 52 elevator pitches were acoustically analyzed by means of PRAAT with respect to the following parameters that reflect the state-of-the-art knowledge about melodic charisma (see [4,12] for an overview of the corresponding findings):

- (1) f0 range (semitones, st): a larger f0 range is more charismatic;
- (2) Lowness of phrase-final f0 (st relative to a speaker's individual baseline f0): lower final falls are more charismatic;
- (3) Net speaking rate (syllables/s, disregarding silent pauses): a faster tempo is more charismatic;
- (4) IPU duration (s): a shorter IPU duration is more charismatic;
- (5) Acoustic-energy level (RMS, decibel, dB): a louder voice is more charismatic.

Measurements were taken automatically based on PRAAT (praat.org) scripts but checked for implausible values, which were then replaced by manual measurements. All measurements were made in increments of 10 ms, and mean values were calculated per IPU. Each elevator-pitch performance and, thus, each speaker is represented by between 33 and 68 IPUs or mean parameter values per condition (before/after training). Note that disfluency phenomena like false starts or fillers were not excluded from the measurements; firstly, because their status for perceived speaker charisma is unclear [19] and, secondly, because they were rare enough throughout all recorded presentations (< 10 on average per speaker) to not bias the analysis.

In addition, all mean parameter values of an elevator pitch were integrated and translated into a total acoustic-melodic charisma (PASCAL) score per speaker. The assessment procedure underlying the PASCAL score was developed on an experimental-phonetic basis. Numerous stimulus series were generated, 50% each with male and female speakers. In one type of series, individual melodic parameters such as speaking rate, f0 level, or f0 range were raised and lowered, for example, with PSOLA resynthesis (http://cnx.org/content/m12474/latest/) in at least 20 equidistant steps. Subsequently, in another type of stimulus series, two melodic parameters were raised or lowered in parallel or in opposite directions. With all stimulus series, online perception experiments were carried out, in which listeners rated the speakers' charisma with values between 0-100. A total of over 500,000 such listener ratings were analyzed over a period of four years. In this way, the gender-specific charisma "sweet spots" per parameter were determined, together with the exact charisma-lowering effect of parameter values outside these "sweet spots". Moreover, a set of numbers (multipliers) was worked out that expresses the relative importance of each parameter in the interplay of all charisma-relevant parameters in a speaker's speech melody.

The PASCAL score quantifies the overall charismatic performance of a speaker as the sum of the parameter-specific performance values, weighted with the individual multipliers per parameter. The score itself is then specified in relation to the performance of the approximately 1,000 speakers whose melodic charisma has so far been assessed by the PASCAL system.

PASCAL scores have repeatedly proven to be reliable and precise in predicting the listener ratings of perceived speaker charisma [20,21]. On this basis, we computed PASCAL scores for all before/after-training presentations in our study and used them in place of formal listener ratings to quantify the 52 speakers' perceived melodic charisma levels and their magnitude of change from before to after training.

3 Results and Discussion

For the inferential statistics, we used two-way repeated-measures analyses of variance (in SPSS v26) with Training (before vs. after) and Software as the two fixed factors and Speaker as a random factor. Two analyses were conducted, one with the five acoustic measurements (MANOVA) and one with the speaker's PASCAL scores (ANOVA) as dependent variable(s). Analyses of variance could be used as previous explorative data analyses, including Kolmogorov-Smirnov and Henze-Zirkler tests, showed that the individual datasets underlying the Training and Software factor levels do not deviate significantly from a normal distribution. Moreover, Greenhouse-Geisser corrections were applied in the analyses of variance, if required.

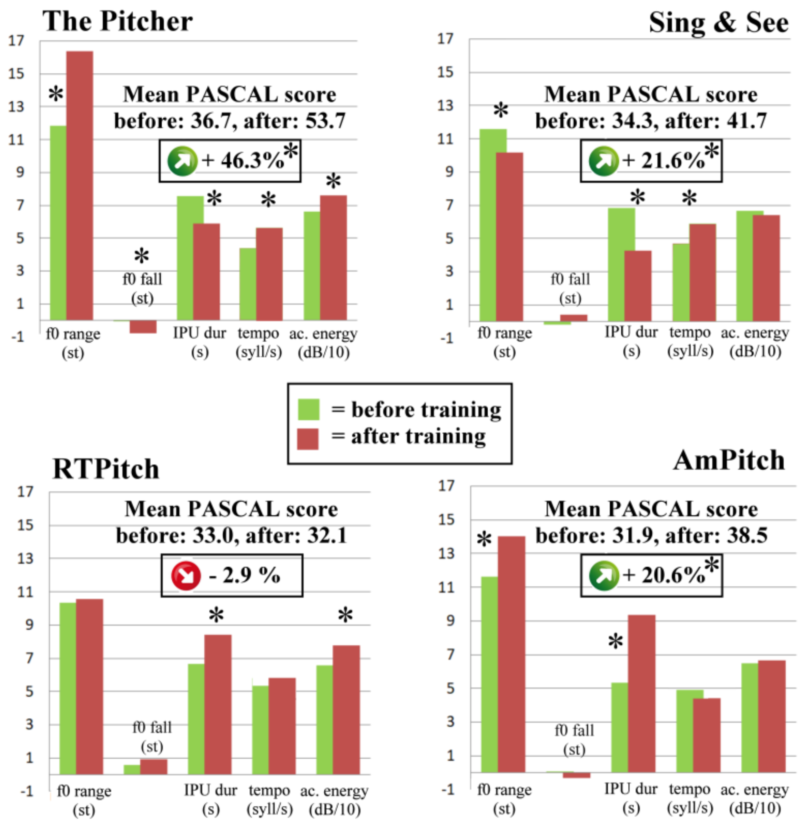

Fig. 4. Mean values for PASCAL scores and the 5 acoustic measures in the 4 Software conditions before/after training. Each value = 13 speakers and about 650 IPUs; <*> = $p<0.05$.

First of all, we found a significant interaction between Training and Software for both the acoustics-based MANOVA ($F[3,408]=53.451$, $p<0.001$, $\eta_p^2=.717$) and the PASCAL-based ANOVA ($F[3,408]=38.992$, $p<0.001$, $\eta_p^2=.663$). The interaction reflects that, while the 4 groups or software-conditions did not differ *before* training, they did differ significantly *after* training. This interaction is a key result as it enables us to safely interpret group differences after training as being caused by the group-specific software tools used for melodic charisma training. Note that the absence of group differences before training does not mean that all speakers performed equally in their initial elevator pitch. It merely means that the students in one group were not generally better or worse than in another group. In fact, the effect of Speaker was significant in both domains acoustics and perception (MANOVA: $F[3,408]= 12.018$, $p<0.001$, $\eta_p^2=.195$; ANOVA: $F[3,408]=16.833$, $p<0.001$, $\eta_p^2=.246$).

Given the significant Training*Software interaction, we split the total sample into the 4 software groups and then separately analyzed the effect of Training for each group. Multiple-comparisons t-tests with Sidak corrections of alpha-error levels were conducted to examine the effects of Training in detail. Only significant differences at $p<0.05$ are reported below with reference to the results summary in Fig. 4.

Regarding the acoustic results, Fig.4 shows that training with the Pitcher had significant charisma-supporting effects along all speech-melody parameters. After training, the f0 range was larger, the phrase-final f0 movements ended lower, the tempo and acoustic-energy levels were higher, and the IPU durations were shorter.

AmPitch was the only tool with a similarly positive effect on f0 as the Pitcher. That the f0 range became significantly larger and the phrase-final fall slightly but significantly lower with AmPitch is probably due to the f0 distribution that gradually emerged on y-axis over the 60 s time window. The downside of being able to see for a longer time period how (much) f0 spreads along the y-axis is the lack of f0-contour resets at short periodic time intervals. This is probably the reason why AmPitch did not motivate speakers to talk faster, and why the IPU duration significantly increased with AmPitch, both of which is very unfavorable for a speaker's charismatic impact.

An unfavorable increase in IPU duration occurred for RTPitch as well whose horizontal-scrolling function also lacked a periodic f0-contour reset. However, while the approach of AmPitch to integrate the acoustic-energy visualization in the thickness of the f0 contour failed to make speakers talk louder after training, the separate waveform-amplitude display of RTPitch was successful in that respect. Like with the dots of the Pitcher, RTPitch users were significantly louder after training. But, besides that RTPitch had no favorable effects on any melodic charisma parameter, including f0.

The charisma-supporting performance of Sing&See was in between those of the other tools. Like for the Pitcher, the Sing&See visualization was sensitive to silent pauses and featured a f0-contour reset at short periodic time intervals; and, like the Pitcher, Sing&See had a charisma-supporting effect on speech timing. That is, it significantly increased the speakers' tempo and reduced their IPU durations after training. In contrast, Sing&See was the only tool that significantly narrowed speakers' f0 range after training and hence reduced their charisma in this respect. Note in this context that Sing&See was also the only tool whose visualized f0 contour was both continuous *and* non-stylized. Also note that, unlike for RTPitch and the Pitcher, the separate acoustic-energy meter of Sing&See had no effect on speakers' loudness level after training, perhaps because the meter is placed too far from the focused f0 contour.

As regards the PASCAL scores that are used here to quantify the speakers' charismatic impact on listeners, Fig. 4 shows a three-part gradation of the software-tools' performances. RTPitch yielded the lowest after-training scores. The speakers even became on average slightly (but not significantly) less charismatic when they trained with this tool. The other three tools all performed better than RTPitch and led to a significant increase in PASCAL scores – and, thus, in the speakers' charismatic impact – from before to after training. Additional t-test comparisons between the after-training samples of the 4 groups showed moreover that Sing&See and AmPitch performed equally well at improving speaker charisma (by 21.6% or 20.6%, respectively), while the Pitcher was still significantly better ($p<0.01$) and outperformed both Sing&See and AmPitch by another 25%, yielding a total increase in speakers' PASCAL scores of on average 46.3% after only a single hour of self-guided training.

4 Conclusion and Outlook

Based on our results, the research questions put forward in 1.2 can be answered as follows: Yes, speakers can learn to use a different, more charismatic speech melody by means of software-supported real-time visualizations of acoustic-melodic features. Yes, the Pitcher does in fact outperform the other visualization tools. This applies to the acoustic measurements (the Pitcher improved all measured parameters and, moreover, proved outstandingly effective for f0-range improvement) as well as to the PASCAL score that quantifies a speaker's charismatic impact on his/her audience and that was used here in place of real listener ratings.

The significantly better charisma-training performance of the Pitcher compared to the other software tools is remarkable, but also not surprising given that the Pitcher was the only one among the 4 compared tools that was specially developed for training a charismatic speech melody. Furthermore, it was the only tool that provided speaker feedback not just in terms of a visualization, but also in terms a color-coded evaluation.

From this point of view, it is striking that the other tools achieved improvements in the speakers' after-training presentation charisma at all. Given how big and software-specific these improvements are, it is unlikely that they stem from a general familiarization effect whose effect on charisma is not consistently positive anyways. Rather, what comes into play here is probably an effect that the authors already know from their coaching experience: The breakdown of a phenomenon as intangible as the speech melody into individual elements (i.e., parameters) such as range, tempo, final fall, etc., in combination with the mere visualization of these elements, gives speakers a meta-understanding of the cause-and-effect relationships for these elements and, thus, a more conscious control over their speech melody. Combined with targets such as "faster", "lower", "louder" etc., this more conscious control can already lead to improvements in a self-guided training task. In this context, the results of our study suggest that the effect size or average improvement potential of this meta-understanding alone is around 20%. For additional evaluative feedback (like the red and green lines and dots in the Pitcher), it is larger and approximately at 30%. An interesting question for future research is to what extent these numbers change relatively and absolutely when the self-guided training session is longer than one hour.

Furthermore, the present results have the following implications for the visualization and learning of speech melody, also with respect to applications beyond charisma training, such as second-language learning: First, all elements of speech melody can be learned and changed through software-based real-time visualizations. Second, in accord with [13] and [14], a continuous but stylized f0 contour is most effective in guiding speakers to specific f0 changes. In contrast, a contour that is both continuous and non-stylized is least effective and even worse than discontinuous, non-stylized contours. Third, an overview of the f0 distribution along the y-axis (as provided by AmPitch) is instructive to learn changes in the f0 range. Fourth, also in accord with [13] and [14], integrated visualizations like that of acoustic-energy and f0 in AmPitch are not effective. Acoustic-energy visualizations (like those of stress and rhythm) must be kept separate from f0 visualizations. Fifth, waveform visualizations are an effective way to provide feedback on acoustic energy or loudness, as are changes in the size and color of syllable-related dots. Sixth, f0-contour resets (both their presence or absence) are effective ways to guide speakers into desired speaking-rate and/or pausing behaviors.

It is up to future studies to examine and substantiate these implications empirically. The authors' next step will be to look at the feedback potential of waveform visualizations, especially for an element of speech melody that plays a role in perceived speaker charisma [22] but is difficult to measure, visualize and learn: voice quality.

References

1. Heide, F.J.: Easy to sense but hard to define: Charismatic nonverbal communication and the psychotherapist. J. of Psychotherapy Integr. 23, pp. 305–319 (2013).
2. Weber, M.: On charisma and institution building. Chicago: Univ. Press (1968).
3. Antonakis, J.; Bastardoz, N.; Jacquart, P.; Shamir, B.: Charisma: An ill-defined and ill-measured gift. Ann. Rev. of Org. Psych. Org. Beh. 3, pp. 293–319 (2016).
4. Michalsky, J.; Niebuhr, O.: Myth busted? Challenging what we think we know about charismatic speech. Acta Univ. Caro. Phil. 2, pp. 27–56 (2019).
5. Chen, L.; Feng, G.; Joe, J.; Leong, C.W.; Kitchen, C.; Lee, C.M.: Towards automated assessment of public speaking skills using multimodal cues. Proc. 16th International Conference on Multimodal Interaction, Istanbul, Turkey (2014).
6. Wörtwein, T.; Chollet, M.; Schauerte, B.; Morency, L.P.; Stiefelhagen, R.; Scherer, S.: Multimodal public speaking performance assessment. Proc. ACM on International Conference on Multimodal Interaction, Seatle, USA, pp. 43–50 (2015).
7. Amon, I.: Die Macht der Stimme. Munich: Redline (2016).
8. Gregory, S.W. Jr.; Gallagher, T.J.: Spectral analysis of candidates' nonverbal vocal communication: predicting U.S. presidential election outcomes. Soc. Psychol. Q. 65, pp 298–308 (2002).
9. Tegtmeier, S.; Schweisfurth, T.; Niebuhr, O.: Gatekeepers' Biases and the Role of Voice in Start-up Pitches. Proc. 1st SEFOS, Sonderborg, Denmark (2019).

10. Fischer, K.; Niebuhr, O.; Jensen, L.C.; Bodenhagen, L.: Speech Melody Matters – How Robots Profit from Using Charismatic Speech. ACM THRI 9, pp. 1–21 (2019).
11. Niebuhr, O.; Michalsky, J.: Computer-generated speaker charisma and its effects on human actions in a car-navigation system experiment. Lec. Notes in Comp. Sc. 11620, pp. 375–390 (2019).
12. Niebuhr, O.; Tegtmeier, S; Schweisfurth, T.: 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).
13. Niebuhr, O.; Alm, M.; Schümchen, N.; Fischer, K: Comparing visualization techniques for learning second language prosody: First results. Int. Journal of Learner Corpus Res. 3, pp. 250–277 (2017).
14. Fischer, K.; Niebuhr, O.; Schümchen, N.: Visualizing Intonation for Foreign Language Teaching. Journal of Lang. Teach. Res. (submitted).
15. Wagner, P.; Origlia, A.; Avesani, C., Christodoulides, G., Cutugno, F., D'Imperio, M. et al.: Different parts of the same elephant: a roadmap to disentangle and connect different perspectives on prosodic prominence. Proc. 18th International Congress of Phonetic Sciences, Glasgow, Scotland, pp. 1-5 (2015).
16. Alderson, J. C.: The CEFR and the need for more research. The Modern Language Journal 91(4), pp. 659-663 (2017).
17. Sondhi, S.; Khan, M.; Vijay, R.; Salhan, A. K.: Vocal indicators of emotional stress. International Journal of Computer Applications 122(15) (2015).
18. Niebuhr, O.; Tegtmeier, S.: Virtual Reality as a Digital Learning Tool in Entrepreneurship: How Virtual Environments Help Entrepreneurs Give More Charismatic Investor Pitches. Digital Entrepreneurship, pp. 123-158 (2019).
19. Niebuhr, O.; Fischer, K.: 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. 20th International Interspeech Conference, Graz, Austria, pp. 544-548 (2019).
20. Niebuhr, O.; Michalsky, J.: PASCAL and DPA: A pilot study on using prosodic competence scores to predict communicative skills for team working and public speaking. Proc. 20th Interspeech Conference, Graz, Austria, pp. 306–310 (2019).
21. Schmidt, N.: Die charismatische Stimme - Der Ton macht die Musik. https://www.daserste.de/information/wissen-kultur/w-wie-wissen/charismatische-Stimme-100.html, last accessed 22/07/2020.
22. Niebuhr, O.; Skarnitzl, R.; & Tylečková, L.: The acoustic fingerprint of a charismatic voice – Initial evidence from correlations between long-term spectral features and listener ratings. Proc. 9th Speech Prosody, Poland, pp. 359-363 (2018).