REVISTA DA **ABRALIN**

RESEARCH REPORT

Computer-assisted prosody training: Improving public speakers' vocal charisma with the Web-Pitcher

Oliver NIEBUHR
Centre for Industrial Electronics – University of Southern Denmark (SDU)

ABSTRACT

Computer-assisted prosody training (CAPT) has so far mainly been used to teach foreign languages, although prosody is still hardly taken into account in language leraning. Conversely, prosody receives a lot of attention in studies and activities related to public-speaker training. But, despite that, CAPT tools are practically unknown in this type of training. The present paper addresses this gap and introduces the "Web Pitcher", a new browser-based version of the feedback and evaluation software "Pitcher" that was developed in 2018 for the prosody-oriented training of charisma – a key characteristic of successful public speakers, which is defined as signaling competence, self-confidence and passion. In an online experiment with 60 test users it is investigated here whether and to what extent the Web Pitcher positively influences the prosodic charisma triggers of its users, and which feedback modes in which order lead to the greatest learning success. An acoustic analysis of before- vs. after-training speeches given by the 60 test users shows that the Web Pitcher improves six key prosodic charisma triggers of its learners by an average of 53 % after one hour of training – and thus performs at eye level with its offline precursor, the Pitcher. With the correct combination of its two feedback modes, the Web Pitcher even outperforms its offline precursor in terms of user improvement. The results are discussed with a view to further R&D steps and the integration of the Web Pitcher in real coaching activities. In this context, the paper also contains a link through which researchers can register and use the Web Pitcher for their own scientific purposes, also beyond questions of public-speaker or charisma training.

OPEN ACCESS

EDITED BY
- Miguel Oliveira Jr. (UFAL)
- Oliver Niebuhr (SDU)

REVIEWED BY
- Plínio Barbosa (UNICAMP)
- Waldemar Ferreira Netto (USP)

DATES
- Received: 01/27/2021
- Accepted: 08/11/2021
- Published: 09/02/2021

HOW TO CITE
Niebuhr, O. (2021). Computer-assisted prosody training: Improving public speakers' vocal charisma with the Web-Pitcher. *Revista da Abralin*, v. 20, n. 1, p. 1-29, 2021.

RESUMEN

El entrenamiento en prosodia asistido por ordenador (CAPT) se ha utilizado hasta ahora principalmente para aprender idiomas extranjeros, aunque la prosodia toda-vía apenas se tiene en cuenta en este campo de investigación y aplicación. Por el contrario, la prosodia recibe mucha atención en los estudios y actividades relacionados con la formación en oratoria. Pero, a pesar de eso, las herramientas CAPT son prácticamente desconocidas en este tipo de formación. El presente estudio aborda esta brecha e introduce el "Web Pitcher", un desarrollo posterior -basado en navegador web- del software de retroalimentación y evaluación "Pitcher" desarrollado en 2018 para el entrenamiento del carisma orientado a la prosodia, una característica clave de los oradores más competentes, que se define como señal de competencia, autoconfianza y pasión (también a través de la prosodia). En un experimento en línea con 60 usuarios de prueba, se investiga si, y en qué medida, el Web Pitcher influye positivamente en los desencadenantes del carisma prosódico de sus usuarios, y qué modos de retroalimentación en qué orden conducen al mayor éxito de aprendizaje. Un análisis acústico de los discursos antes y después del entrenamiento realizado por los 60 usuarios de prueba muestra que el Web Pitcher mejora seis factores desencadenantes clave del carisma prosódico de sus alumnos en un promedio del 53% después de una hora de entrenamiento y, por lo tanto, funciona al mismo nivel que su precursor oficial, el Pitcher. Con la combinación correcta de sus dos modos de retroalimentación, Web Pitcher incluso supera a su precursor offline en términos de mejora del usuario. Los resultados se discuten con vistas a nuevos pasos de I+D y la integración del Web Pitcher en actividades reales de coaching. En este contexto, el artículo también contiene un enlace a través del cual los investigadores pueden registrarse y utilizar el Web Pitcher para sus propios fines científicos, más allá de las cuestiones de formación en oratoria o carisma.

KEYWORDS

Charisma. Prosody. Public speaking. Web Pitcher. Pitch. Tempo. Pausing. CAPT. English.

PALABRAS CLAVE

Charisma. Prosodia. Oratoria. Web Pitcher. Tono. Ritmo. Pausa. CAPT. Inglés.

1. Introduction

This article is about computer-assisted pronunciation training or, more precisely, computer-assisted prosody training. It is the latter training that we will refer to as CAPT, following the line of research of Yenkimaleki & Van Heuven (2019). On the empirical basis of a proof-of-concept test, we present the first stage of development of a new browser-based prosodic-feedback tool that we share with the scientific community for research purposes on request. The corresponding registration form can be viewed and downloaded here[1].

1.1 CAPT: So far a matter of foreign-language learning

CAPT represents a newly emerging field of research and development activities. As a branch of the relatively well-researched and long-standing computer-assisted language learning (CALL, see TAFAZOLI et al., 2019), the aim of CAPT is to help speakers learn the prosody of a new language with software support and on the basis of automatic, individual user feedback.

Prosody is a complex bundle of stress, loudness, intonation, and voice quality, cf. Arvaniti (2021). The four types of phenomena are coordinated with remarkable precision by speakers to create meaningful and functional prosodic patterns. Probably all the world's languages use prosody in order to (a) encode emotional states and speaker attitudes towards the interlocutor or the exchanged information, and in order to (b) control the interaction dynamics and hierarchy between speaker and listener, see Mozziconacci (2001) or Kohler (1991). In many languages, the signaling of word accents, the syntagmatic structuring of utterances, and the marking of sentence mode and information structure also belong to the inventory of prosodic functions (HEDBERG; SOSA, 2008). Research shows that prosodic elements are the first ones we learn in our mother tongue – in fact, prenatally already, see Mampe et al. (2009) and Langus et al. (2017).

Gilbert (2008) argues that, for learners of a foreign language or dialect, new vowel sounds are harder to acquire than new consonant sounds, because, unlike consonants, vowels do not provide learners with the tactile feedback that they can use as landmarks for readjusting their articulation (e.g., in the form of tongue contacts with the palate, teeth, or lips). When we consider that what is true of vowels must be all the more true of prosody – and when we additionally consider that prosody lacks, for naïve speakers, the tangible and syntagmatically delimitable ways in which vowels and consonants form meaningful words, then it becomes clear why prosody is not just the first thing we learn in our own language, but also the last thing we can put down from our own language when we switch to a nonnative language or dialect (MENNEN; DE LEEUW, 2014).

This is where CAPT tools come into play. With a focus on foreign languages rather than dialects, the tools provide learner feedback, for example, in that they visualize the pitch contour of the foreign-

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

language learner on a screen in real time relative to a reference contour of the language to be learned (DEMENKO et al., 2009; SZTAHÓ et al., 2018; PYSHKIN et al., 2019). This applies to the "Tell me more" application of Rosettastone (LIAW; ENGLISH, 2017) as well as to the iOS mobile-phone application "StudyIntonation" (LEZHENIN et al., 2017; PYSHKIN et al., 2019) and the tool "Nordplus" for Scandinavian languages by Fischer et al. (2021). Only a few tools currently go beyond intonation. The tool of Su et al. (2018) records an utterance made by the foreign language learner and then plays a computer-generated (resynthesized) version of this utterance, but with a corrected, native-like prosody. This basically allows the user to not only acquire the respective intonational form-function link, but also the coinciding stress or rhythm patterns of the foreign-language utterances. Additionally, s/he gets insights into how expressive and emotional prosodies work in the target language and interact with the linguistic form-function links of prosody, see also Bonneau et al. (2004) for a similar approach.

Independently of what feedback the respective CAPT tool provides to the foreign-language learner and when, all tools share the same pedagogical approach: they support the foreign-language learner in that they contrast the learner's own prosodic realization with that of the target language – be it in the form of a visual or an auditory stimulus. In addition, CAPT tools usually operate on the basis of individual short utterances and specific prosodic functions such as the indication of sentence mode, focus, or lexical tone. Pyshkin et al. (2019) point out in this context that CAPT tools are likely to work more effectively if the training of the respective prosodic function is embedded in a specific and well-substantiated (e.g., multi-medially illustrated) conversational context. Similarly, the relevance of social context was stressed by study of Amrate (2021) who showed that their CAPT tool performed better when learners practiced with it not alone but in pairs.

1.2 The relevance of CAPT beyond foreign-language learning

The key point of departure of this paper is that no language has only a *single* prosody. Every language additionally possesses registers, i.e., specific linguistic 'settings' that are used for precisely defined purposes or communicative situations, cf. Agha (2004) and Prsir et al. (2014). These settings are, of course, also partly morphosyntactically and grammatically defined. But, above all, they are defined prosodically. Reading news in the media is such a setting (COTTER, 1993; BARBOSA et al., 2017), as well as praying or preaching in church (NIEBUHR; SCHJOEDT, 2019; FELEVE; FAJOBI, 2019), telling a story (WICHMANN, 2021; BARBOSA et al., 2017), talking to babies/children, i.e., motherese (BIERSACK et al., 2017) – or giving a speech (ATKINSON, 2004; SOORJOO, 2012).

The present paper deals with the setting of the latter register, i.e., the prosody of public speaking. In contrast to a foreign-language prosody, it can be assumed that speakers in principle know how to create the prosodic settings for the registers of their own language. In connection with public speaking, however, some disruptive factors come into effect that cause great prosodic differences between speakers: public-speaking anxiety (HSU 2009), personality traits (MICHALSKY et al., 2020), the availability and nature of audience feedback (CHOLLET et al., 2015), or the difference between

read speech on the one hand and spontaneous speech on the other (MIXDORFF; PFITZINGER, 2005), with the latter being the desired output and the former being the typical starting point of a public speech (e.g., in the form of text on presentation slides). In short: "Public speaking is an art, but at the same time a skill" (POP; CRISAN, 2014, p. 44); and, as Pop and Crisan further emphasize: "[...] a skill which is formed through constant practice" (p. 44). Similarly, Gehrke (2016, p. 261) concludes in his manifesto for teaching public speaking that "public speaking is an art, and like so many arts, robust scientific research can offer some guidance on how to teach it".

In view of such statements, of which there are many similar ones, the following fact is remarkable. In foreign-language learning, prosody is strongly marginalized in favor of lexicon and grammar, mainly due to the teachers' lack of prosodic knowledge and lack of prosodic teaching material (UKAM et al., 2017; LEVIS, 2018; FISCHER et al. 2021). Nevertheless, there is already a significant number of CAPT tools to support users in learning a foreign language prosody, see 1.1. In contrast, practically every public-speaking manual or trainer devotes at least one main chapter or coaching session to the topic of prosody (ATKINSON, 2004; SOORJOO, 2012; MORTENSEN, 2011). Research also shows that prosodic elements play a key role for the perceived persuasiveness, self-confidence, and passion – in short the charisma – of a public speaker (WÖRTWEIN et al., 2015; CHEN et al., 2014). So, unlike in foreign-language learning, prosody is anything but marginalized in public-speaker training; and yet, the entire field of training manages without the use of CAPT tools and instead falls back on traditional, vague, descriptive instructions such as "use an animated voice" or "speak fluently" (ATKINSON, 2004; SOORJOO, 2012; MORTENSEN, 2011).

To fill this gap, we have developed a new prosody-oriented learning tool for public-speaker training: the Pitcher (NIEBUHR; NEITSCH, 2020). The name does not (primarily) refer to the perceived melody of speech, but rather to a particular type of short presentation that people give to present their ideas, products, or plans to a jury or an audience of peers or experts. There are different subtypes of such 'pitches', known as business pitch, elevator pitch, investor pitch, etc. (SABAJ et al., 2020).

The Pitcher is embedded in an overall concept for digital rhetorical training[2]. Note in connection with this concept that the pedagogical framework for such a CAPT tool cannot simply be adopted from foreign-language learning. In contrast to foreign-language learning (cf. 1.1), public-speaker training is about mastering *holistic* prosodic *settings*, not about mastering *local* prosodic *patterns*. Prosodic learning can therefore neither take place in relation to a specific reference prosody nor on the basis of short individual sentences. Furthermore, CAPT tools for public-speaker training cannot focus on specific communicative functions such as sentence mode or focus; and while CAPT tools for foreign-language learning can to some extent be limited to the pitch contour, a CAPT tool for public speaker training must be broader and provide user feedback also for tempo, pausing, loudness, etc.

[2] For further information see also https://www.allgoodspeakers.com/training

There is great potential for the use of CAPT tools in public-speaker training. The Pitcher is to help develop this new area of application and research. The Pitcher visualizes in real time prosodic key parameters of charismatic public speaking and also evaluates these parameters in real time using an integrative metric – PICSA – that was derived from about 500,000 listener ratings of prosodically manipulated speech stimuli (NIEBUHR; NEITSCH, 2020). At the end of a speech, the Pitcher summarizes the prosodic contribution to the public-speaking performance and provides the learner with an overall charisma score. The listener ratings for PICSA came from participants with a Western-Germanic language background, i.e., Dutch, German, English. Therefore, PICSA works best for these languages.

The present paper presents a further developed version of the Pitcher. Unlike the original Pitcher, this further developed version is browser-based and, thus, comes a little closer to our main goal of offering everyone the possibility of mobile, effective, and entertaining public-speaker training; a training that focuses on the key rhetorical element of the speaker's voice. The performance of the original Pitcher with regard to teaching its users a charismatic public-speaking prosody was tested in a speech-production experiment by Niebuhr and Neitsch (2020). Our goal here is, firstly, to replicate this experiment and its positive results with the further developed browser-based Pitcher (henceforth the Web Pitcher) and, secondly, to evaluate the new functionalities of the Web Pitcher in order to make recommendations for an effective use of these new functionalities in rhetorical practice. In addition, the paper aims to introduce the Web Pitcher to the scientific community. As mentioned above, we make this CAPT tool available to researchers for scientific purposes.

2. The (Web) Pitcher software

The CAPT tool 'Pitcher' was developed to help speakers learn the essential prosodic ingredients of a charismatic vocal delivery in business and sales presentations or other (e.g., political) public-speaking scenarios. Charisma is defined here with reference to Michalsky and Niebuhr (2019) as the combined signaling (through various visual and acoustic means, including prosody) of competence, self-confidence and passion. The definition is inspired by Antonakis et al. (2016) who defined charisma as "values-based, symbolic, and emotion-laden leader signaling" (p. 17). Compared to the latter definition, the definition of Michalsky and Niebuhr (2019) is more specific on the "leader" element and, at the same time, it takes up the tripartite structure of the PAD model (pleasure, arousal, dominance) from environmental or consumer psychology (DONOVAN; ROSSITER, 1982; NAGANO et al., 2021).

There is increasing evidence for cross-linguistic differences in both the production and the perception of a charismatic vocal delivery (BIADSY et al., 2008; D'ERRICO et al., 2013; GUTNYK et al., 2021), which is one reason why the present paper is, like that of Niebuhr and Neitsch (2020), restricted to public-speaking performances in Western Germanic languages or, more specifically, to (nonnative) English. However, the interfering influence of the factor language on perceived speaker charisma is mitigated insofar as cross-linguistic differences in the production and perception of charisma are of a quantitative rather than qualitative nature. That is, they refer to the lower

threshold above which a prosodic parameter setting begins to unfold a charismatic effect and/or the upper limit above which it starts to sound inauthentic or exaggerated and has a negative effect on the speaker's charisma. There are hardly any qualitative cross-linguistic differences, for example, in the sense of prosodic parameters that (a) are relevant in language A but irrelevant in language B or that (b) correlate in opposite directions with perceived speaker charisma in A and B. That cross-linguistic differences are quantitative rather than qualitative in nature makes sense given that the prosodic signaling of competence, self-confidence, and passion correlates with the universal prosodic form-function links of the 'biological codes'. This applies in particular to the frequency code, the effort code, and the size code (GUSSENHOVEN, 2016; XU et al., 2013). We can therefore assume that while the Pitcher's feedback settings have to be adjusted cross-linguistically and, in particular, beyond Western Germanic languages, the prosodic parameters for which feedback is given as well as the underlying feedback concepts of the Pitcher are universally applicable for all languages.

The Pitcher measures the level, range, and variability of the pitch in a learner's voice as well as how low s/he is able to fall with the pitch, e.g., at the end of a concluding statement (henceforth referred to as 'final fall'). Moreover, the frequency of occurrence of silent pauses as well as their duration are measured by the Pitcher, as are the speaking rate (tempo) and the loudness level and variability of the learner's speech.

For each measurement, the Pitcher offers simple color-coded feedback, with green, yellow, and red indicating whether the learner is, for each prosodic feature, inside, in the periphery, or outside of the parametric window that supports perceived speaker charisma. This window is also referred to as the 'sweet spot'[3]. The location of these parametric sweet spots along each prosodic feature are, amongst other things, speaker- and gender-specific. Therefore, the Pitcher requires that the user calibrates the Pitcher with his/her specific voice prior to getting color-coded real-time feedback on his/her public-speaking performance. This calibration happens in the original Pitcher by means of the open vowel [a], which the speaker is required to produce for 3-5 seconds at a comfortable pitch level.

We now turn to the further developed Web Pitcher. It differs in several respects from the original Pitcher. For example, the user-specific calibration procedure was simplified such that the Web Pitcher asks the user to assign his/her voice-pitch level to a specific kind of cat, i.e., baby cat, house cat, bobcat, tiger, or lion. This further development was for one thing implemented as a measure of political correctness. It avoids that users have to specify their biological sex or gender before they can use the Web Pitcher. This specification has only limited information value in any case, as the voice-pitch levels of male and female speakers are not categorically distinct, see, for example, Carullo et al. (2013). For another thing, selecting the individual voice-pitch level with reference to cat species of different sizes has the additional advantage that it avoids a calibration of the Web Pitcher to the speaker's voice over the Internet, which can be a quite error-prone and noisy

[3] "a location or combination of characteristics that produces the best results", https://www.oxfordlearnersdictionaries.com/definition/english/sweet-spot?q=sweet+spot)

procedure when sound equipment and mouth-to-microphone distances are unknown and the room acoustics is not ideal.

Finally, the Web Pitcher also allows the user to choose between three levels of difficulty: Beginner, Advanced Learner, and Expert, see Fig. 1. These levels determine how strictly the public-speaking performance is assessed by the system. That is, the selected level of difficulty narrows or widens the parametric size of each sweet spot.

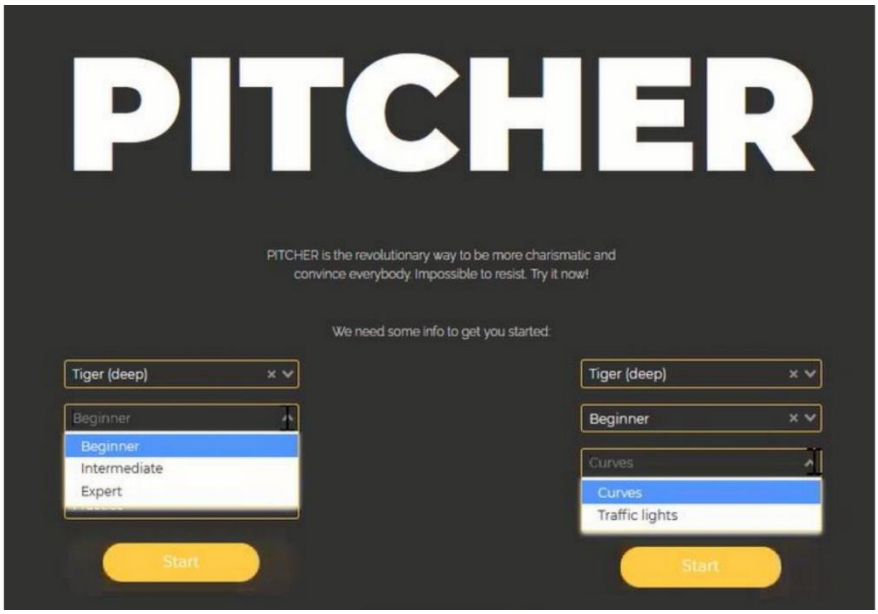


FIGURE 1 – Edited screenshot of the Web-Pitcher's main-menu display.
Source: the author.

Another new feature of the Web Pitcher is that the user can choose between two different feedback modes: the 'curve mode' and the 'traffic-light mode', see Figure 1 and Figure 2a-b. The original Pitcher only offered the curve mode.

The curve mode works in real time. As is shown in Figure 2a, it draws the pitch contour as a stylized line, based on a smoothing technique (Note that the curve mode actually draws not perceived pitch itself, but its acoustic pendant, i.e., the fundamental frequency or f0 in Hz; however, to reach a large readership with this paper, we rely on simple terms here and continue to use pitch or intonation to refer to f0 patterns). Stylization was found to be critical to successful training as it reduces the amount of melodic detail to a level that is manageable and instructive for learners, see Fischer et al. (2021). The stylized intonation contour is overlaid with dots that mark the individual spoken syllables. As the learner's tempo increases, the color of the dots changes from green to yellow; and if the tempo gets too fast the dots turn red, indicating to the user that s/he needs to slow down until the dots turn green again. The upper and lower horizontal lines mark the minimum pitch range that the learner needs to cover or, ideally, exceed with his/her intonation in order to be perceived as charismatic. Thus, the learner has to rise with the voice pitch above the upper line and fall

below the lower line (occasionally and when it is appropriate in terms of intonational functions) to be within the sweet spot of the pitch range. Crossing the lower line is also what matters for a charismatic final fall. That is, it is a characteristic of charismatic speakers that they end their statements with a fall until the bottom of their pitch range. Approximately in the middle of the two pitch-range lines there is another horizontal auxiliary line. It shows where the learner's optimal, charismatic pitch level is located, i.e., the level around which his/her rising/falling pitch movements have to vary. All three lines are red by default and change color from yellow to green, if the user crosses them often enough with his/her intonation or if, in the case of the middle line, the average pitch (measured over the previous 5 seconds) is within the sweet spot of the pitch level feature.

The intonation contour is drawn from left to right across the screen along a time axis of 5 seconds. This time interval is chosen because it represents the minimum rate of silent pauses in public speaking. That is, the user should at least pause once in the course of his/her speech before the right end of the screen is reached to be within the sweet spot of this prosodic feature. For pauses with the required minimum duration of 500 ms the timeline is reset, and the intonation-contour drawing starts again from the left end of the screen. In addition, the color with which the intonation contour is drawn changes from blue to red just before the right end of the screen is reached. The color change reminds the user that a silent pause (> 500 ms) is overdue. Note in this context that the Web Pitcher also measures the duration of these silent pauses > 500 ms. However, so far at least, there is no visual and evaluative (i.e., color-coded) feedback for pause duration.

In traffic-light mode, which did not exist in the original Pitcher, each prosodic feature is represented by a traffic light that updates every 5 seconds. Green means that the speaker has on average been within the sweet-spot window of the respective parameter in the past 5 seconds. Red means the opposite and, thus, prompts the speaker to improve in that respective parameter in the course of the following speech.

It is on purpose that the traffic-light mode does not work in real time. Receiving simple and binary (red vs. green, i.e., good vs. bad) feedback at longer time intervals reduces the feedback density for the learner and enables him/her to concentrate better on individual prosodic features. Moreover, the traffic-light mode is overall less distracting, for example, for speakers who would like to get a general overview of their performance during a video or phone call with a customer or business partner. In addition, a performance assessment, which abstracts from local strengths and weaknesses over a larger time window, is closer to the actual behavior of the listener when it comes to perceiving speaker charisma (TSKHAY et al., 2017; CASPI et al., 2019).

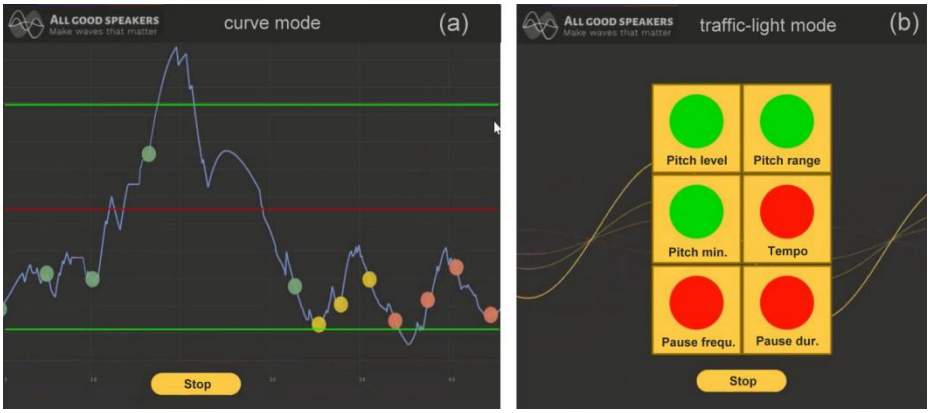


FIGURE 2 – Edited screenshots of the Web-Pitcher's (a) curve mode and (b) traffic-light mode.
Source: the author.

When the user ends the feedback session, an evaluative summary appears on the screen in the same way for both the curve and the traffic-light mode. It specifies the percentage of the total speaking time during which the user was within the sweet spot of each prosodic feature, see Figure 3. The feedback list includes pause duration, although there is currently no visual and evaluative feedback for this prosodic feature in the curve and traffic-light modes. However, it was assumed that users would find this pause-duration feedback still useful and learn from it for their next round of practice.

As a small gamification add-on, the summary screen also displays the image and name of that celebrity who comes closest to the learner's performance on each individual prosodic feature. Clicking on one of these images directs the user to a YouTube video in which s/he can hear what it sounds like when the celebrity uses the respective prosodic feature in the same way as the user did. The Web Pitcher includes a database with the averaged public-speaking performances of over 200 celebrities – from artists like Madonna, Justin Bieber, Brad Pitt and George Clooney, through politicians like Angela Merkel, Boris Johnson, Queen Elisabeth II, and Benjamin Netanyahu to CEOs like Mark Zuckerberg, Elon Musk, Jan Hsun Huang, and Zhang Yong. The database is gender-balanced and grows successively.

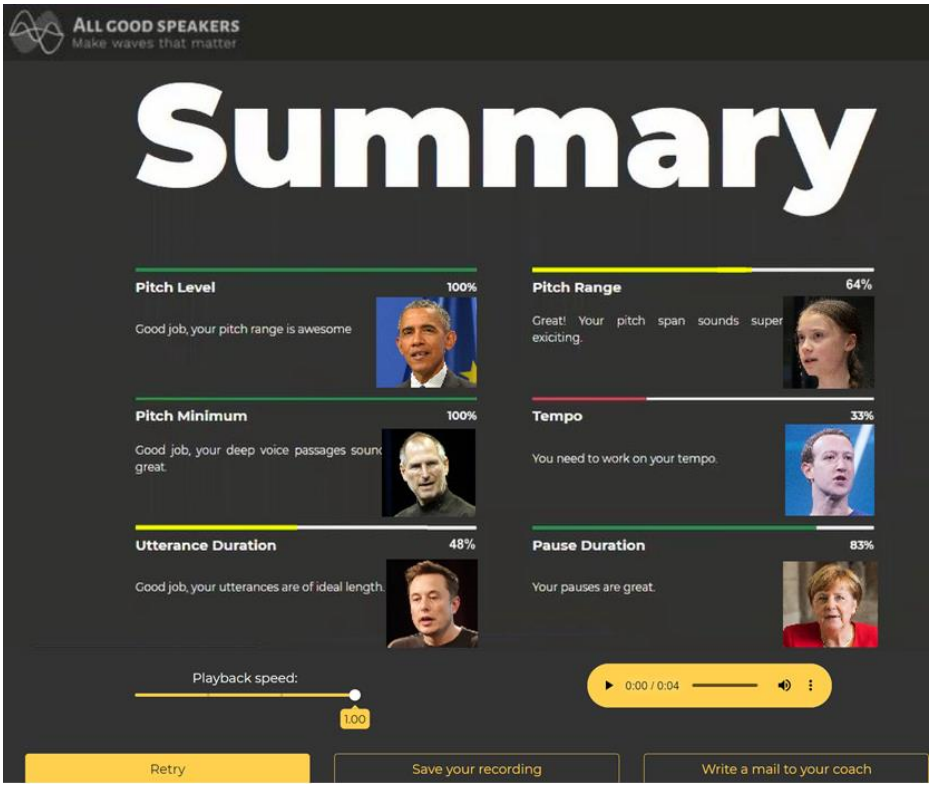


FIGURE 3 – Screenshot of the Web-Pitcher's summary display following each feedback session.
Source: the author.

Furthermore, the users have the option to play their own presentation and listen to what they did and how (at different playback speeds) as well as to save their recording and/or send their recording along with questions and comments directly to their public-speaking trainer.

3. How prosodic features are correlated with speaker charisma

Next to the phonetic sciences (ROSENBERG; HIRSCHBERG, 2009; NIEBUHR et al., 2016; BOSKER, 2020), other disciplines such as psychology (MILLER et al., 1976; GREGORY; GALLAGHER, 2002), economics (DAVIS et al., 2017) and speech technology (FISCHER et al., 2019) also investigated the relationship between prosodic features and the charismatic impact of speakers. Evidence from empirical and, in particular, experimental research characterizes this relationship as highly complex. For one thing, this is because the three key traits associated with speaker charisma – i.e., competence, self-confidence, and passion – stand in a complex interplay with prosodic features. Since manuals of rhetoric do not sufficiently consider or penetrate this interplay, they sometimes come up with deviating or contradictory recommendations for learners of public speaking. For example,

some manuals recommend raising the voice pitch level for public speaking, whereas other manuals recommend lowering it. Closer scientific analysis help resolve such supposed contradictions. In the case of the voice pitch level, for example, research shows that both recommendations are correct, but apply to different points in the intonation contour (MICHALSKY; NIEBUHR, 2019) and in combination with different loudness levels.

Another reason why the relationship between prosodic features and the charismatic impact of public speakers is complex is that the sweet spots vary depending on gender and, therefore, may require men and women to deviate from the baseline levels of their prosodic parameters in different directions. Baseline levels refer to those levels that characterize the speakers' speech in everyday, matter-of-fact contexts (i.e., outside contexts in which speech is "performed" like in presentations). For example, see the reference values that are reported in Andreeva et al. (2014), Pépiot (2014), and Volín et al. (2015). Moreover, the sweet spots are limited at *both* ends. This means that there is not only a threshold above which the realization of a prosodic feature unfolds an increasingly positive charisma effect. There is also a threshold above which a further parametric change in the same direction increasingly damages the speaker's charisma (NIEBUHR; NEITSCH, 2020, cf. also ROSENBERG; HIRSCHBERG, 2009). The latter is rarely referred to in rhetorical manuals, cf. Atkinson (2004), Soorjoo (2012), or Mortensen (2010). For example, Mortensen (2010) underlines in his rhetorical manual that "people who speak faster appear more competent and knowledgeable" (p. 156). Accordingly, Mortensen recommends his readers to increase the speaking rate to enhance their charisma in presentations. However, he does not say how much faster (than the everyday matter-of-fact baseline) is actually advantageous for speakers and at what speed 'fast' becomes 'too fast'. It has long been known that clear speech and fast speech are to a certain extent incompatible (BERNSTEIN et al., 1992), and that it is clear speech rather than strongly reduced and mumbled speech that supports the charismatic impact of a speaker (NIEBUHR; GONZALEZ, 2019). Thus, there must be an upper limit for the increase of a speaker's speech rate in presentations.

In the present paper, however, we are only interested in whether the Web Pitcher software is able to shift the prosodic feature values of its users in a direction that is typically beneficial with respect to average baseline values (e.g., towards more frequent silent pauses); and/or whether it can prevent its users from shifting their prosodic feature values in a non-beneficial direction, again with respect to average baseline values (e.g., towards a too narrow and hence too monotonous-sounding pitch range). On this basis, we can simplify the complex relationships between prosody and charisma and assume, for the purposes of the present study, unidirectional correlations between prosodic feature values and perceived speaker charisma. The following applies in detail:

- Level, range, and variability of the pitch in a learner's voice are positively correlated with perceived speaker charisma. That is, the higher-pitched the voice is and the more often and extensively the pitch is varied, the more charismatic the speaker sounds.

- The final fall is negatively correlated with perceived speaker charisma. That is, the lower a speaker is able to fall with the pitch of his/her voice at the ends of utterances (like concluding statements), the more charismatic s/he sounds.

- Tempo is negatively correlated with perceived speaker charisma. The slower a speech is produced, the more charismatic the speaker sounds.

- The frequency of occurrence of silent pauses as well as their duration are both positively correlated with perceived speaker charisma. That is, the more often the speaker pauses and the longer his/her pauses are, the more charismatic s/he sounds.

Note that, unlike in the experimental test of the original Pitcher by Niebuhr and Neitsch (2020), the Web Pitcher omits the analysis and evaluation of loudness features. This is because analyzing and, in particular, evaluating loudness becomes an unreliable matter if, for example, the learner's mouth-to-microphone distance as well as the frequency response curve of the microphone are unknown or vary depending on where and when the learner uses the Web Pitcher. The original Pitcher measures loudness in terms of RMS intensity (dB). Of course, there are alternative measures of loudness, such as 'spectral emphasis'. Heldner (2003) stresses that "overall intensity and spectral emphasis [...] represent two different operationalizations of loudness" and that the latter is a more robust measure than the former. However, simply switching from intensity to spectral emphasis for the purposes of the present study is not possible for two reasons. Firstly, the Web Pitcher currently does not calculate spectral emphasis, i.e., these measurements are unavailable and, secondly, while the quantitative relationship between intensity (RMS) and perceived speaker charisma is known and represented in the Pitcher's underlying charisma-evaluation metric PICSA, this does – to date – not apply to spectral emphasis. Thus, before measurements and user feedback for spectral emphasis can be implemented in the Web Pitcher, at least the gender-specific sweet spots of this measure need to be defined.

Note further that pitch variability is not the same as pitch range. A larger pitch range and a lower final fall make the pitch variability increase as well. However, what primarily shapes the variability measure is how often the speaker's pitch goes up and down, which, in turn, reflects how many words s/he stresses in his/her speech. Measuring this frequency of stressed words is the main purpose of the variability measure. (Actually, stressed words in the given context mean pitch-accented words. However, for the same reason why we refer to f0 as pitch, we will stick to 'stress' as the simpler and less theory- or field-specific term, cf. LADD, 2008).

4. Research questions

The following four questions are addressed:

I. Can we replicate the positive learning effects obtained for the original Pitcher with the Web Pitcher? In other words, after training with the Web Pitcher, do learners produce a larger pitch range, make more frequent silent pauses, and use a slower tempo and a lower final fall?

II. Beyond what was measured with the original Pitcher, do learners after using the Web Pitcher additionally show a higher voice pitch level and longer silent pauses (although the latter is the only measured parameter without visual, evaluative feedback)?

III. Is learning in the traffic-light mode as effective as in the curve mode?

IV. If both modes are used by learners subsequently, which order is more effective, curve mode followed by traffic-light mode or traffic-light mode followed by curve mode?

At the beginning of the experiment of Niebuhr and Neitsch (2020) that we aim to replicate here, all participants took part in an introductory lecture in perceived speaker charisma and its prosodic triggers. The lecture was followed by a short public speech of each participant. The speeches were held and recorded in a classroom setting (the university's largest lecture hall) in front of an audience of peers. Afterwards, the participants were asked to practice what they had heard in the lecture individually and independently with the Pitcher for one hour. Then, the same short public speeches were held and recorded a second time under the same conditions. Based on that, the prosodic performance in the speakers' before-practice (baseline) speeches was compared with that in their after-practice (test) speeches. We will also apply this paradigm in the present study, but with the difference that the experiment was moved to the Internet. This was not because COVID-19 restrictions were in force at the time the experiment was carried out, but mainly because the Internet corresponds to the real application environment of the Web Pitcher. All further details of the method are described below.

5. Experiment

5.1 Participants

The experiment was conducted with a total of 60 participants, 29 women and 31 men. They were between 22-46 years old, recruited via online (social-media) platforms and paid for their participation. All participants were naive insofar as they had no prior training in, or in-depth experience with public speaking tasks (beyond occasional presentations given in the course of their education). Moreover, all participants were fluent nonnative speakers of English and reported no speech or hearing disorders.

The fact that we analyze nonnative instead of native English is in principle relevant, but negligible for the present study because the question here is how an audience with an English background (i.e., in the social and linguistic interpretation mode 'English') would rate the vocal charisma of the speakers' presentations. As Levis (2018) points out: "many, if not most, interactions in English around the world take place without the involvement of a native speaker" (p.3). In view of this and with reference to the fact that the Pitcher and Web Pitcher are both mainly used as CAPT tools in nonnative English business contexts, our nonnative speaker sample is appropriately chosen and ecologically valid.

5.2 Procedure

About 3-5 days before their first (baseline) speech, the 60 participants received a link to a pre-recorded e-learning video about perceived speaker charisma and its prosodic triggers. The video explained what speaker charisma is and how relevant it is for professional and private life according to research. Furthermore, the concept of prosody was introduced, and its central role for speaker charisma was described. Finally, the video contained concrete instructions about which prosodic changes have a positive or negative effect on perceived speaker charisma. The email to the participants also contained a short (approx. 2-3 minutes long) pre-formulated product presentation or, more specifically, an investor pitch, cf. Sabaj et al. (2020). The participants were instructed to familiarize themselves with this text so that they could perform the presentation fluently in their baseline speech a few days later.

The before-practice (baseline) performance of the given product presentation was then recorded with all 60 participants in individual meetings. The performance of the presentation as well as its recording took place in the environment of a Zoom call. The participants took part in the Zoom call without video transmission in order to increase the audio quality and avoid the risk of strong compression of the audio signal in favor of the video signal. Moreover, participants were asked to minimize the reverb suppression of Zoom and switch on the Windows audio drivers in the Zoom settings. Zoom uses the OPUS codec for speech-signal processing, and with the applied audio

settings, this codec has little or no effects on measurements of prosodic charisma features, including pitch, see Siegert and Niebuhr (2021).

On the screen shared by the experimenter, the participants saw a video excerpt from a real multi-party Zoom meeting played in an endless loop during the presentation recording[4]. This was to make the presentation situation as realistic as possible for the participants, see Figure 4.

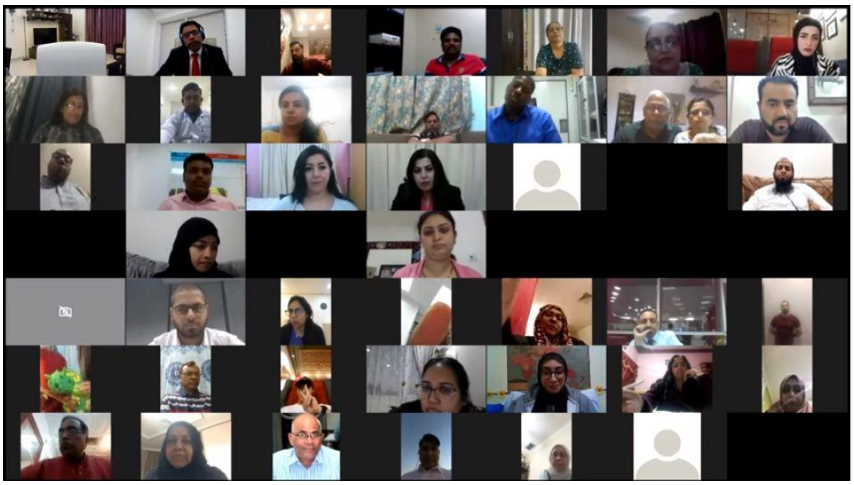


FIGURE 4 – Screenshot of the edited video excerpt played to the participants in order to create more realistic public-speaking/presentation conditions. The excerpt was repeated in an endless loop.
Source: the author, edited after https://www.youtube.com/watch?v=Jmw00iPlPKI.

Subsequently to the baseline recording, the 60 participants were divided up and semi-randomly assigned to 4 equally large groups of 15 participants. The assignment was semi-random as participant gender was balanced in each group. All 4 groups received via email a link and a password with which they could access the Web Pitcher software. What differed between the groups was the instruction included in the email text. Group 1 received the instruction to use only the Web Pitcher's curve mode to practice a more charismatic presentation prosody. Group 2, by contrast, was asked to practice only in the traffic-light mode. The other two groups were told to use both modes for half the practice time each – Group 3 in the sequence of curve mode and traffic-light mode, and Group 4 in the sequence of traffic-light mode and curve mode.

The entire practice time with the Web Pitcher was the same for all groups, i.e., one hour. This means that Groups 3 and 4 were asked to practice their presentations with the curve and the traffic-light mode for about 30 minutes each. In all groups, participants set the level of difficulty of the Web-Pitcher's feedback to 'Beginner'.

Immediately after the participants reported back (by email) to the experimenter that they had finished their Web-Pitcher practice hour (no participant fell significantly short of this time interval), the after-practice (test) presentation of the short product presentation was given and recorded. The

[4] The video excerpt was created based on https://www.youtube.com/watch?v=Jmw00iPlPKI

presentation and recording conditions were identical to those of the baseline round. That is, once again the participants were recorded in individual audio-only sessions, and they gave their presentation in front of the same multi-party Zoom audience illustrated in Fig. 4.

All 2 x 60 = 120 baseline and test recordings were made while the participants and the experimenter sat in silent rooms in their own home.

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. They mean less stress and, thus, a lower pitch level (SONDHI et al., 2015), whereas a higher pitch level is required for charismatic speech (ROSENBERG; HIRSCHBERG, 2009; NIEBUHR et al., 2016). Moreover, intensive presentation training quickly causes routine/boredom, which also has unfavorable effects on the prosodic aspects of speaker charisma (NIEBUHR; TEGTMEIER, 2019). Finally, and most importantly, comparing the four different feedback-mode conditions is a between-subjects matter. Thus, any potential within-subjects familiarization effects have no influence on the results of these between-subject comparisons (as there is no reason to assume that the magnitude of this familiarization effect differs systematically between the semi-randomly compiled groups).

5.3 Acoustic analysis

The participants' recorded baseline and test presentations were acoustically analyzed by means of PRAAT with respect to all prosodic features listed in the correlations of section 3. Pitch measurements were made in semitones (st) relative to a reference level of 100 Hz (men) or 200 Hz (women). Tempo was measured in syllables per second (syll/s). Duration measurements were made in seconds (s). The pause frequency was also expressed as a duration value, more precisely as the duration of the units of speech in between two pauses. These units are known as inter-pausal units or IPUs. For a speech whose total duration is predetermined by a constant, given text like the investor pitch used here, IPU durations become shorter the more pauses the speaker inserts into the presentation.

Measurements were carried out automatically with scripts in PRAAT (praat.org) and checked for implausible values, which were then either omitted or replaced by manual measurements. The measurement procedure was based on the IPUs in the participants' presentations. That is, one value per IPU was calculated. In the case of pitch features, which were measured in increments of 10 ms across an IPU, the mean values per IPU were used. The participants' presentations included between 39 and 55 IPUs each.

Disfluency phenomena like false starts or fillers were not excluded from the measurements; firstly, because their status for perceived speaker charisma is unclear (NIEBUHR; FISCHER, 2019) and, secondly, because due to the pre-formulated text and the long familiarization/preparation

phase they were rare enough throughout all recorded presentations (< 7 on average per speaker) to not bias the analysis.

5.4 Results

The results of the acoustic analysis are summarized in Fig. 5 and Table 1. Table 1 shows the test statistics of a series of two-way ANOVAs, run with the fixed factors Recording (baseline vs. test) and Group (the four feedback-mode conditions) for each of the 7 independent prosodic-feature variables. The measurements of the 15 participants per group were pooled, creating sample sizes between n = 589 (Group 1) and n = 703 (Group 3). Because of these relatively large sample sizes and taking into account multiple testing, the alpha error level (i.e., the significance threshold) was set to $p < 0.01$.

Pros. feature	**Main effects Rec and Group**	**Interactions Rec x Group**
Pitch level (mean f0, st)	Recording: F = 196.79, p < 0.001, $\eta_p^2 = 0.55$ Group: F = 45.04, p < 0.001, $\eta_p^2 = 0.31$	F= 59.03, p< 0.001, η_p^2= 0.27
Pitch range (f0 range, st)	Recording: F = 2324.46, p< 0.001, η_p^2= 0.92 Group: F = 445.78, p < 0.001, $\eta_p^2 = 0.78$	F=448.40, p<0.001, η_p^2= 0.73
Pitch variability (f0 sd, st)	Recording: F = 71.80, p < 0.001, $\eta_p^2 = 0.37$ Group: F = 13.18, p < 0.001, $\eta_p^2 = 0.29$	F = 15.74, p<0.001, η_p^2= 0.22
Final fall (f0 min, st)	Recording: F = 674.85, p < 0.001, $\eta_p^2 = 0.73$ Group: F = 126.38, p < 0.001, $\eta_p^2 = 0.65$	F= 126.63, p<0.001, η_p^2=0.48
Tempo (syll/s)	Recording: F = 100.29, p < 0.001, $\eta_p^2 = 0.49$ Group: F = 467.99, p < 0.001, $\eta_p^2 = 0.66$	F= 113.68, p<0.001, η_p^2=0.40
Pause frequ./ IPU duration (s)	Recording: F = 1061.05, p< 0.001, η_p^2= 0.84 Group: F = 208.36, p < 0.001, $\eta_p^2 = 0.60$	F=232.71, p<0.001, η_p^2= 0.59
Pause duration (s)	Recording: F = 0.23, n.s. Group: F = 0.44, n.s.	F = 0.65, n.s.

TABLE 1 – Test statistics of the conducted two-way ANOVAs; *df1* = 1 (Group) or *df1* = 3 (Recording and Interaction Rec x Group); *df2* = 5,204 in all cases; η_p^2 values indicate effect sizes.
Source: the author.

Table 1 shows that all ANOVAs yielded significant main effects of Recording and Group as well as a significant Recording x Group interaction. One exception was the prosodic feature pause duration, which was the only feature the Web Pitcher provided no direct visualization and color-coded feedback for.

Tukey's HSD tests were used to shed light on the sources of the significant main effects and interactions. A uniform pattern emerged across all 7 features: Firstly, in the baseline recordings there were no significant differences between Groups 1-4 for any of the 7 analyzed prosodic features. Of course, there was inter-speaker variance in all groups, but the four groups as a whole behaved statistically identically in their prosodic implementation of the baseline presentation. Secondly, this prosodic baseline implementation clearly differed from that in the later test recording. This difference caused the main effects of Recording per prosodic feature. Thirdly, the prosodic differences

between baseline and test recording were not equivalent across the four groups. Some groups showed greater prosodic changes from baseline to test recording than others. The significant interactions Recording x Group rely on this group-specific effect of Web-Pitcher training on speech prosody, together with the fact that significant group differences were limited to the test recordings.

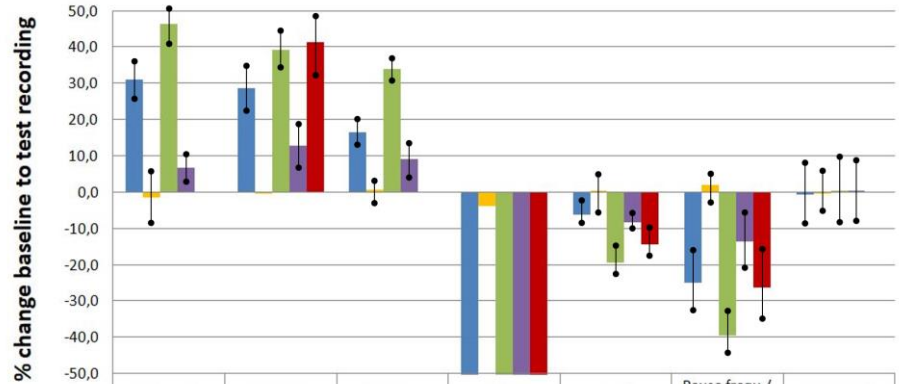
Figure 5 illustrates the results in detail. In order to integrate all prosodic features into one bar graph despite different scales and acoustic parameter levels, we display the change in % from baseline to test recording per group. Note that some %-bars of the final fall exceed the percentage scale on the y-axis. The scale was not adjusted, however, for the sake of a good resolution of the other percentage differences. The exact %-change values of the truncated bars as well as of all other bars are specified in the table below the graph.

Firstly, Figure 5 shows that, in terms of the relative magnitude of the percentage changes, the Web Pitcher feedback had a stronger impact on the pitch features than on the tempo and duration features. Secondly, there was a group-specific magnitude of change. For example, while Group 1 members, who practiced in the curve mode only, showed a clear change from baseline to test recording, there was practically no such change (also no statistically significant one) for Group 2 members, who practiced in the traffic-light mode only. The prosody of Group 3 members changed the most. They first practiced for 30 minutes in the curve mode and then for 30 minutes in the traffic-light mode. If this feedback-mode sequence is reversed, as in Group 4, significant prosodic changes still occurred. However, these were significantly smaller, not only compared to those of Group 3 members, but also compared to those of Group 1 members, who practiced in the curve mode for the entire 60 minutes. Note in this context that all changes (in terms of group averages) went in the direction of an enhanced speaker charisma and were consistent with how the color-coded feedback concept of the Web Pitcher was intended to support the learner.

Thirdly, Figure 5 also shows the relative baseline-to-test changes that emerged for the experimental testing of the original Pitcher by Niebuhr and Neitsch (2020). Note that only 4 of the 7 prosodic features could be compared to those of the Web Pitcher as the original Pitcher did not measure and evaluate pitch level, pitch variability, and pause duration. It can clearly be seen that the results obtained for the original Pitcher by Niebuhr and Neitsch were replicated in the present study, mainly in terms of direction of change, but to some degree also with respect to the relative magnitude of change. Regarding the latter, for example, we see that the percentage changes (decreases) in tempo were smaller than the changes (increases) in pitch range, which, in turn, were smaller than the changes (decreases) in final fall.

Beyond these general similarities, we also see a complex pattern of differences between the percentage changes obtained by the original Pitcher and the four different groups (i.e., feedback conditions) of the Web Pitcher, with Group 1 being the one whose curve-mode-only training most closely matched the training in the original Pitcher experiment. For pitch range and tempo, separate t-tests show that the changes triggered by the original Pitcher were significantly larger than those of Web-Pitcher Group 1 ($t[1299] = 5.44$, $p < 0.001$, $d = 0.67$), but on par with those of Web-Pitcher Group 3. In the case of pause frequency or IPU duration, Group 1 of the Web Pitcher performed at

eye level with that of the original Pitcher, whereas Group 3 of the Web Pitcher performed significantly better ($t[1313] = 6.21$, $p < 0.001$, $d = 0.75$). Finally, both Web Pitcher Groups 1 and 3 were able to outperform the original Pitcher group in the extent to which they lowered their final pitch fall ($t[1286_{Group1}/1245_{Group3}] > 8.55$, $p < 0.001$, $d > 0.88$).



	Pitch level	Pitch range	Pitch varia.	Final fall	Tempo	Pause frequ./ IPU duration	Pause dur
Group1	31,0	28,5	16,4	**-171,0 (16,5)**	-6,3	-25,0	-0,7
Group2	-1,5	-0,3	0,6	**-3,8 (4,7)**	0,3	1,8	-0,3
Group3	46,3	39,1	33,9	**-272,5 (28,8)**	-19,5	-39,6	0,3
Group4	6,6	12,8	9,2	**-98,0 (20,3)**	-8,4	-13,6	0,0
Orig. Pitcher		41,2		**-123,3 (18,9)**	-14,5	-26,4	

FIGURE 5 – Bar graph of the relative changes in speech prosody from baseline to test recording for Groups 1-4 ($589 < n < 703$ per bar). Below the graph are the corresponding % values; vertical lines indicate standard deviations.
Source: the author.

Finally, note with regard to Table 1 that the largest effect sizes (η_p^2) were found for pitch range and pause frequency (or IPU duration), whereas the smallest effect sizes resulted for pitch level and pitch variability. This does not mean that pitch range and final fall have changed the most and pitch level and pitch variability the least as a function of Recording and Group. Rather, what it means is that the variability included in the measurements of pitch range and final fall could more successfully be explained by Recording and Group and their interaction than for pitch level and pitch variability. Recall that Recording and Group represent different prosodic feedback and learning conditions. Thus, the different effect sizes show that pitch range and final fall are more easily influenced by prosodic feedback and learning conditions and less susceptible to unconditioned inter- and intra-individual variability than pitch level and pitch variability. Both is in accord with our own experience from public-speaker training. However, it would be premature to conclude that pitch level and pitch variability are generally less accessible to CAPT approaches. Maybe, we just have not found the right feedback strategies as yet to train pitch level and pitch variability more effectively.

6. Discussion

The conducted experiment was to address three questions. (I) Can we replicate the positive learning effects of the original Pitcher with the Web Pitcher? In other words, after training with the Web Pitcher, do learners produce a larger pitch range, make more frequent silent pauses, and use a slower tempo and a lower final fall? The clear answer to question (I) is YES; and unlike in the experiment on the original Pitcher by Niebuhr and Neitsch (2020), we tested not just a single user group in the present experiment, but four user groups. The fact that the prosodic realizations of the (semi-)randomly compiled Groups 1-4 only differed between the test recordings but *not* between the baseline recordings underpins that the prosodic differences of the test recordings are not a simple consequence of intrinsically different group skills that already existed prior to the learning phase. Rather, we conclude that the prosodic differences between Groups 1-4 in the test recordings represent genuine learning effects created by working with the Web Pitcher in different ways.

Of course, any prosodic changes from baseline to test recording could in principle also represent training artifacts of familiarity or boredom. An explanation of the found changes through such artifacts is, however, unlikely; firstly, because the overall pattern of changes from baseline to test recordings does not match known prosodic artifacts of familiarization or boredom (e.g., all pitch features should show decreases, not increases, see NIEBUHR; TEGTMEIER, 2019); and secondly, because the differences between Groups 1-4 in the test recordings are too heterogeneous to be due to uniform artifacts of familiarization or boredom.

Question (II) asked if learners, after using the Web Pitcher, additionally show longer silent pauses and a higher voice-pitch level. The obtained evidence does not provide a uniform answer to question (II). YES, the Web Pitcher was able to raise the average voice-pitch level of its users significantly, especially for Groups 1 and 3. In the experiment with the original Pitcher, the voice-pitch level was not measured. Informal comparisons between these earlier data and the new Web-Pitcher data suggest, however, that training with the original Pitcher also raised the voice pitch level, albeit not as strongly as with the Web Pitcher and probably more as an indirect consequence of an increase in loudness, see Watson and Hughes (2006) and references therein for further details on the correlation of pitch and loudness. It can therefore be assumed that the reference line for a charismatic voice-pitch level, which Web Pitcher offered as a new feature compared to the original Pitcher, was effective in improving the users' performance on this prosodic feature. The second part of the answer to question (II) is NO. We did not find any evidence of longer silent pauses as a result of Web-Pitcher training. Note that the prosodic feature 'silent-pause duration' was not just the only feature for which the test recordings showed no significant change relative to the baseline recordings. It was also the only feature for which the Web Pitcher offered no direct visualization and color-coded feedback. In this respect, the missing effects for silent-pause duration indirectly support the effectiveness of the visualization and feedback concepts of the Web Pitcher for the other prosodic features – and, moreover, they support the conclusion that the prosodic changes in the test recordings are genuine learning effects rather than artifacts of familiarization and boredom.

Questions (III)-(IV) concerned the feedback modes themselves. Question (III) asked whether learning in the traffic-light mode would be as effective as in the curve mode. The clear empirical answer to this question is NO. Our results even suggest that practicing presentations in the traffic-light mode alone does not result in any learning effect at all. In contrast to this, the users who only had the curve mode available showed clear learning effects towards a more charismatic presentation prosody. However, the traffic-light mode is not a useless mode either. Group 3, which used the traffic-light mode after the curve mode for practice, showed significantly greater success in learning a more charismatic prosody than Group 1, which only practiced its presentation prosody with the curve mode – and for half an hour longer than Group 3.

In this context, the answer to question (IV) becomes relevant: YES, it actually made a considerable difference in which order the users received the two feedback modes for practicing. Group 4 (traffic-light mode followed by curve mode) not only performed significantly worse than Group 3 (curve mode followed by traffic-light mode), but also worse than Group 1 (curve mode only). This shows that the combination of the two feedback modes is, per se, no benefit for the user. It seems to be decisive – and agrees with our coaching experience – that the immediate experience of an interactive and prosodically concrete real-time feedback as in the curve mode must come *before* the prosodically and temporally more abstract and visually reduced feedback of the traffic-light mode. When practicing with the curve mode, according to our interpretation of the findings, users learn the important links between the prosodic motor commands of speech production on the one hand and the prosodic target patterns of their own voice on the other. In other words, we assume that the curve mode acts as a mediator between the otherwise intangible (in terms of tactile feedback) production of speech prosody and its perception. Only when these links between production and perception are learned can the traffic-light mode, with its abstract and reduced feedback, help consolidate the learning and implementation of a more charismatic prosody.

A last discussion point concerns the Web Pitcher's performance in relation to that of the original Pitcher. We showed that the users of the original Pitcher improved significantly more in some features of a charismatic prosody than Group 1 of the Web Pitcher (curve mode only). This is especially true for pitch range and tempo. One possible explanation for this could be that the participant group of the original Pitcher happened to contain more adaptive individuals than Group 1 of the Web Pitcher; 15 or 13 participants in the respective experimental groups allow only a limited generalization of the findings, so that this explanation cannot be ruled out. A negative effect of e-learning is the more likely explanation, though. In the experiment with the original Pitcher, participants received a real classroom lecture given by a professional charismatic-speech coach. The baseline and test recordings also took place in front of a real audience in a large lecture hall. The high levels of attention and motivation of the participants in such a setting can hardly be replicated in an online setup, with the worse performance of the Web Pitcher Group 1 being the logical consequence.

Based on this explanation, it is all the more important to emphasize that Group 3 of the Web Pitcher (curve mode followed by traffic-light mode) performed considerably better in the baseline-to-test comparison than the group of the original Pitcher. Web Pitcher Group 3 achieved

similarly large charismatic improvements for pitch range and tempo, and clearly outperformed the original Pitcher group in terms of a deeper final fall and a higher pause frequency (meaning shorter IPU durations). In concrete figures, the original Pitcher improved its users by 51.35 % on average across all measured prosodic-feature changes, while Group 3 of the Web Pitcher improved by an average of 93.42 %. This underlines the relative success of the tested Web Pitcher, also and especially for learning a charismatic prosody in an online setup. However, to put this relatively high percentage increase into perspective, one has to take three things into account: First, the participants were students whose only experience with public speaking relates to oral examinations and presentations as part of their education. It is likely that business professionals with more experience in public speaking already perform at a higher baseline level and, correspondingly, show smaller percentage improvements in a Web-Pitcher training. Second, prosodic public speaker training and the corresponding Web Pitcher feedback are about holistic changes. Thus, what users of the Web Pitcher are supposed to learn is in some way similar to "singing" a song at a different tempo or on a different pitch. It can be assumed that such prosodic changes are easier to learn than the complex patterns of stress, pitch, and their coordination that participants are confronted with in foreign-language learning. Therefore, the percentage improvement obtainable through CAPT tools in prosodic public-speaker training is probably inherently higher than in foreign language learning. Third, nothing is currently known about the sustainability of these improvements. That is, even if the members of Group 3 were able to change their prosodic settings by almost 94% on average, that does not mean that this change is sustainable. However, if used on a regular basis and with occasional supervision, mobile and easily accessible CAPT tools seem to offer an ideal basis for achieving this kind of sustainability.

Conclusion and Outlook

With regard to computer-assisted prosody training (CAPT), our experiment provided compelling evidence that successful and largely self-directed online training of a more charismatic vocal performance is feasible with new, science-based tools such as the Web Pitcher. This encompasses all the prosodic core features of this vocal performance, such as pitch, tempo, and pausing. Furthermore, regarding practical coaching, our results show that the traffic-light mode in itself is not a useful feedback mode, but can significantly increase learning performance when used as an additional feedback mode after curve-mode training. Why exactly this is so and whether the provided explanation of learning (curve mode) and consolidating (traffic-light mode) the links between the production and perception of charismatic prosody patterns is correct has to be addressed in follow-up studies. It is also important to increase the number of participants in follow-up studies in order to achieve a more solid generalization of the conclusions made. The evidence obtained here encourages these follow-up studies to take the next step and move from controlled experimental settings to testing the Web Pitcher under real "field" conditions.

Finally, against the background of the available evidence, it is also worthwhile to further expand and refine the visualization and feedback concepts of the Web Pitcher. This applies, for example, to the feature of silent-pause duration. New visualization and feedback concepts could also be developed for the pitch variability feature. In assessing vocal charisma, this feature is not only relevant as an integrative measure of pitch range and final fall, but also the indicator of the number of stressed words in a presentation (up to a certain limit, a larger number of stressed words is considered positive for perceived speakers charisma). The feedback for pitch level can possibly also be made more effective. The present experiment suggests that evaluative feedback in the form of a color-coded reference line indicating the most charismatic voice-pitch level (for the given speaker and gender) already had a beneficial effect. The relatively low effect size and the inflexibility of speakers to change their voice pitch level known from practical experience, however, suggest that there are better ways of providing visual feedback on the pitch level.

In addition to the step-by-step improvement of feedback strategies, the Web Pitcher can also be supplemented with other prosodic features relevant to vocal charisma. This includes in particular the level and variability of the loudness of the speaker's voice. Against the background of Heldner (2003) and similar studies, we will use 'spectral emphasis' for this and adopt the successful loudness-visualization strategy from the original Pitcher: a change in the size of the dots of the individual syllables proportional to their loudness level, combined with a color change of the dots to red if the loudness level falls permanently below a lower loudness threshold.

All these plans, tests, and tasks will also stimulate new fruitful interdisciplinary collaborations between experts in the fields of phonetics, design, digital speech signal processing, and computer linguistics. As was stated in 1.2: Apart from the Pitcher and the Web Pitcher, there are currently to the best of our knowledge no other CAPT tools in the field of voice-oriented public-speaker training. At the same time, there is great potential in this field for the use of such tools, not least because the voice plays such a fundamental role in the perceived charisma of a speaker. Against this background, we are happy to make the Web Pitcher available for research purposes on request, and we hope for a lively transfer of knowledge and results.

Acknowledgements

The author is greatly indebted to his two reviewers for their useful and insightful comments on an earlier draft of this paper. Moreover, thanks are due to Merikan Koyun, Christian Lüke, and Stephan Senkbeil for their commitment and suggestions in creating the Web Pitcher. A last special "thank you" goes to Io Valls for translating the abstract of the paper into Spanish. Finally, note that the author is also the CEO of the speech-technology company AllGoodSpeakers ApS. Please visit https://www.allgoodspeakers.com/coi for a corresponding conflict-of-interest statement.

REFERENCES

AGHA, A. Registers of language. A companion to linguistic anthropology, v. 23, p. 45, 2004.

AMRATE, M. Collaborative vs. individual computer-assisted prosody training: a mixed-method case study with Algerian EFL undergraduates. *Computer Assisted Language Learning*, p. 1-28, 2021.

ANDREEVA, B.; DEMENKO, G.; WOLSKA, M.; MÖBIUS, B.; ZIMMERER, F.; JÜGLER, J.; TROUVAIN, J. Comparison of Pitch Range and Pitch Variation in Slavic and Germanic Languages. In: Proc. 7th International Conference on Speech Prosody, Dublin, Ireland, 2014, p. 776-780.

ARVANITI, A. The phonetics of prosody. In: Aronoff, M. Oxford Research Encyclopedia of Linguistics. Oxford: Oxford University Press, 2021.

ATKINSON, M. Lend me your ears - All you need to know about making speeches and presentations. Random House, Chatham, 2004.

BARBOSA, P. A.; MADUREIRA, S.; DE MAREÜIL, P. B. Cross-Linguistic Distinctions Between Professional and Non-Professional Speaking Styles. In: Proc. 18th International Interspeech Conference, Stockholm, Sweden, 2017, p. 3921-3925.

BIADSY, F.; ROSENBERG, A.; CARLSON, R.; HIRSCHBERG, J.; STRANGERT, E. A cross-cultural comparison of American, Palestinian, and Swedish perception of charismatic speech. In: Proc. 4th International Conference of Speech Prosody, Campinas, Brazil, 2008, p. 579-82.

BIERSACK, S.; KEMPE, V.; KNAPTON, L. Fine-tuning speech registers: a comparison of the prosodic features of child-directed and foreigner-directed speech. In: Proc. 9th European Conference on Speech Communication and Technology, Lisbon, Portugal, 2005, p. 1-4.

BONNEAU, A.; CAMUS, M.; LAPRIE, Y.; COLOTTE, V. A computer-assisted learning of English prosody for French students. In: InSTIL/ICALL Symposium on Computer Assisted Learning, Venice, Italy, 2004, p. 1-4.

BOSKER, H. R. The Contribution of Amplitude Modulations in Speech to Perceived Charisma. In: Trouvain, J.; Weiss, B.; Barkat-Defradas, M.; Ohala, J.J. Voice Attractiveness. Singagore, Springer, 2020, p. 165-181.

CARULLO, A.; VALLAN, A.; ASTOLFI, A. Design issues for a portable vocal analyzer. IEEE Transactions on instrumentation and measurement, v. 62, p. 1084-1093, 2013.

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, p. 1-5.

CHOLLET, M.; WÖRTWEIN, T.; MORENCY, L. P.; SHAPIRO, A.; SCHERER, S. Exploring feedback strategies to improve public speaking: an interactive virtual audience framework. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Osaka, Japan, 2015, p. 1143-1154.

COTTER, C. Prosodic aspects of broadcast news register. Annual Meeting of the Berkeley Linguistics Society, v. 19, p. 90-100, 1993.

DAVIS, B.C; HMIELESKI, K.M; WEBB, J.W; COOMBS, J.E. Funders' positive affective reactions to entrepreneurs' crowdfunding pitches: The influence of perceived product creativity and entrepreneurial passion. Journal of Business Venturing, v. 32, p. 90–106, 2017.

DEMENKO, G.; WAGNER, A.; CYLWIK, N.; JOKISCH, O. An audiovisual feedback system for acquiring L2 pronunciation and L2 prosody. In: International Workshop on Speech and Language Technology in Education, Wroxall Abbey Estate, Warwickshire, England, 2009, p. 113-116.

D'ERRICO, F.; SIGNORELLO, R.; DEMOLIN, D.; POGGI, I. The perception of charisma from voice: A cross-cultural study. In: IEEE Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, Switzerland, 2013, p. 552-557.

FALEYE, J. O.; FAJOBI, E. O.. A Prosodic Analysis of English Sermons of Selected Pastors in Southwest Nigeria. Language, v. 3, p. 1-23, 2019.

FISCHER, K.; NIEBUHR, O.; JENSEN, L.C.; BODENHAGEN, L. Speech Melody Matters – How Robots Profit from Using Charismatic Speech. ACM Transactions in Human Robot Interactions, v. 9, p. 1–21, 2019.

FISCHER, K.; NIEBUHR, O.; ALM, M.; SCHÜMCHEN, N. Intuitive Visualization of Intonation for Foreign Language Learners. Submitted.

GILBERT, J. B. Teaching pronunciation. Cambridge, Cambridge University Press, 2008.

GREGORY, S. W. Jr.; GALLAGHER, T. J. Spectral analysis of candidates' nonverbal vocal communication: predicting U.S. presidential election outcomes. Soc. Psychol. Q., v. 65, p. 298–308, 2002.

GUSSENHOVEN, C. Foundations of intonational meaning: Anatomical and physiological factors. Topics in Cognitive Science, v. 8, p. 425-434, 2016.

GUTNYK, A.; NIEBUHR, O.; GU, W. Speaker charisma analyzed through the cultural lens. In: Proc. 12th IEEE International Symposium on Chinese Spoken Language Processing (ISCSLP), Shanghai, China, 2021, p. 1-5.

HEDBERG, N.; SOSA, J.M. The Prosody of Topic and Focus in Spontaneous English Dialogue. In: Lee C.; Gordon M.; Büring D. Topic and Focus. Studies in Linguistics and Philosophy. Springer; Dordrecht, 2008, p. 101-120.

HELDNER, M. On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. Journal of Phonetics, v. 31, p. 39-62, 2003.

HSU, C. F. The relationships of trait anxiety, audience nonverbal feedback, and attributions to public speaking state anxiety. *Communication Research Reports*, v. 26, p. 237-246, 2009.

HUANG, B. H.; JUN, S.-A. The Effect of Age on the Acquisition of Second Language Prosody. Language and Speech, v. 54, p. 387–414, 2011.

KOHLER, K.J. Terminal intonation patterns in single-accent utterances of German: phonetics, phonology and semantics. Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK), v. 25, p. 15-185, 1991.

LADD, D. R. Intonational phonology. Cambridge: Cambridge University Press, 2008.

LANGUS, A.; MEHLER, J.; NESPOR, M. Rhythm in language acquisition. Neuroscience & Biobehavioral Reviews, v. 81, p. 158-166, 2017.

LEVIS, J. M.. Intelligibility, oral communication, and the teaching of pronunciation. Cambridge, Cambridge University Press, , 2018.

LEZHENIN, Y.; LAMTEV, A.; DYACHKOV, V.; BOITSOVA, E.; VYLEGZHANINA, K.; BOGACH, N. Study intonation: Mobile environment for prosody teaching. In: 3rd IEEE International Conference on Cybernetics (CYBCONF), Exeter, UK, 2017, p. 1-2.

LIAW, M. L.; ENGLISH, K. Technologies for teaching and learning L2 reading. In: Chapelle, C.A.; Sauro, S. *The handbook of technology and second language teaching and learning*. London: Wiley & Sons, 2017, p. 62-76.

MAMPE, B.; FRIEDERICI, A. D.; CHRISTOPHE, A.; WERMKE, K. Newborns' cry melody is shaped by their native language. Current biology, v. 19, p. 1994-1997, 2009.

MENNEN, I.; DE LEEUW, E. Beyond segments: Prosody in SLA. Studies in Second Language Acquisition, v. 36, p. 183-194, 2014.

MICHALSKY, J., NIEBUHR, O.: Myth busted? Challenging what we think we know about charismatic speech. Acta Univ. Caro. Phil., v. 2, p. 27–56, 2019.

MILLER, N.; MARUYAMA, G.; BEABER, R. J.; VALONE, K. Speed of speech and persuasion. Journal of Personality and Social Psychology, v. 34, p. 615–624, 1976.

MIXDORFF, H.; PFITZINGER, H. R. Analysing fundamental frequency contours and local speech rate in map task dialogs. *Speech Communication*, v. 46, p. 310-325, 2005.

MORTENSEN, K. W.. The laws of charisma: How to captivate, inspire, and influence for maximum success. Amacom Books, New York, 2010.

MOZZICONACCI, S.. Emotion and attitude conveyed in speech by means of prosody. In: Proc. 2nd Workshop on Attitude, Personality and Emotions in User-Adapted Interaction, Sonthofen, Germany, 2001, p. 1-10.

NIEBUHR, O., VOßE, J.; BREM, A. What makes a charismatic speaker? A computer-based acoustic-prosodic analysis of Steve Jobs tone of voice. Computers in Human Behavior v. 64, p. 366-382, 2016.

NIEBUHR, O.; GONZALEZ, S. Do sound segments contribute to sounding charismatic? Evidence from a case study of Steve Jobs' and Mark Zuckerberg's vowel spaces. International Journal of Acoustics and Vibration, v. 24, p. 343-355, 2019.

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. In: Proc. 20th International Interspeech Conference, Graz, Austria, 2019, p. 544–548.

NIEBUHR, O.; SCHJOEDT, U. God as Interlocutor-Real or Imaginary? Prosodic Markers of Dialogue Speech and Expected Efficacy in Spoken Prayer. In: Proc. 20th International Interspeech Conference, Graz, Austria, 2019, p. 36-40.

NIEBUHR, O.; TEGTMEIER, S. Virtual reality as a digital learning tool in entrepreneurship: how virtual environments help entrepreneurs give more charismatic investor pitches. In: Baierl, R.; Behrens, J.; Brem, A. Digital Entrepreneurship. Springer, Cham, 2019, p. 123–158.

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, v. 4, 12, 2019.

NIEBUHR, O.; NEITSCH, J. Digital Rhetoric 2.0: How to Train Charismatic Speaking with Speech-Melody Visualization Software. Lecture Notes in Computer Science, v. 12335, p. 357-368, 2020.

PÉPIOT, E. Male and female speech: a study of mean f0, f0 range, phonation type and speech rate in Parisian French and American English speakers. In: Proc. 7th International Conference of Speech Prosody, Dublin, Ireland, 2014, p. 305-309.

PRSIR, T.; GOLDMAN, J. P.; AUCHLIN, A. Prosodic features of situational variation across nine speaking styles in French. Journal of Speech Sciences, v. 4, p. 41-60, 2014.

PYSHKIN, E.; BLAKE, J.; LAMTEV, A.; LEZHENIN, I.; ZHUIKOV, A.; BOGACH, N. Prosody training mobile application: Early design assessment and lessons learned. In: Proc. 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, Metz, France, 2019, p. 735-740.

ROSENBERG, A.; HIRSCHBERG, J. Charisma perception from text and speech. Speech Communication, v. 51, p. 640-655, 2009.

SABAJ, O.; CABEZAS, P.; VARAS, G.; GONZÁLEZ-VERGARA, C.; PINA-STRANGER, Á. Empirical literature on the business pitch: Classes, critiques and future trends. Journal of technology management & innovation, v. 15, p. 55-63, 2020.

SIEGERT, I.; NIEBUHR, O. Case Report: Women, Be Aware that Your Vocal Charisma can Dwindle in Remote Meetings. Front. Commun. 5: 611555, p. 1-8, 2021.

SOORJOO, M. Here's the Pitch: How to Pitch Your Business to Anyone, Get Funded, and Win Clients. John Wiley & Sons, London, 2012.

SONDHI, S.; KHAN, M.; VIJAY, R.; SALHAN, A.K. Vocal indicators of emotional stress. Int. J. Comput. Appl., v. 122, p. 38-43, 2015.

SU, C. Y.; TSENG, C. Y.; JANG, J. S. R.; VISCEGLIA, T. A hierarchical linguistic information-based model of English prosody: L2 data analysis and implications for computer-assisted language learning. Computer Speech & Language, v. 51, p. 44-67, 2018.

SZTAHÓ, D.; KISS, G.; VICSI, K. Computer based speech prosody teaching system. Computer Speech & Language, v. 50, p. 126-140, 2018.

TAFAZOLI, D.; HUERTAS ABRIL, C. A.; GÓMEZ PARRA, M. E. Technology-based review on Computer-Assisted Language Learning: A chronological perspective. Pixel-Bit: Revista de Medios y Educación, v. 54, p. 29-43, 2019.

TSKHAY, K. O.; ZHU, R.; RULE, N. O. Perceptions of charisma from thin slices of behavior predict leadership prototypicality judgments. The Leadership Quarterly, v. 28, p. 555-562, 2017.

UKAM, E. I.; UWEN, G. O.; OMALE, C. Application of stress, rhythm and intonation in the speech of Erei-English bilinguals. Global Journal of Arts, Humanities and Social Sciences, v. 5, p. 27-38, 2017.

VOLÍN, J.; POESOVÁ, K.; WEINGARTOVÁ, L. Speech melody properties in English, Czech and Czech English: Reference and interference. *Research in Language*, v. 13, p. 107-123, 2015.

WATSON, P. J.; HUGHES, D. The relationship of vocal loudness manipulation to prosodic F0 and durational variables in healthy adults. Journal of Speech, Language, and Hearing Research, v. 49, p. 636-644, 2006.

WICHMANN, A. Reading aloud: the role of the reader and the conception of 'self'. Journal of Interdisciplinary Voice Studies, 2021.

WÖRTWEIN, T.; CHOLLET, M.; SCHAUERTE, B.; MORENCY, L. P.; STIEFELHAGEN, R.; SCHERER, S. Multimodal public speaking performance assessment. In Proc. 2015 ACM on International Conference on Multimodal Interaction, Seatle, USA, 2015, p. 43-50.

XU, Y.; LEE, A.; WU, W. L.; LIU, X.; BIRKHOLZ, P. Human vocal attractiveness as signaled by body size projection. PloS one, 8(4), e62397, 2013.

YENKIMALEKI, M.; VAN HEUVEN, V. J. The relative contribution of computer assisted prosody training vs. instructor based prosody teaching in developing speaking skills by interpreter trainees: An experimental study. Speech Communication v. 107, p. 48-57, 2019.