How to pause in charismatic speeches: A case study of Barack Obama

Elina Banzina[1], *Oliver Niebuhr*[2]

[1]Dept. of Language and Communication, Stockholm School of Economics, Riga, Latvia
[2]Centre for Industrial Electronics, University of Southern Denmark, Sønderborg/DK
elina.banzina@sseriga.edu, olni@sdu.dk

ABSTRACT

Pauses are well-known to augment speakers' persuasiveness and charisma, yet their effective implementation may go far beyond the standard technique of pausing frequently, prolonging the pauses, and distributing them skillfully. This case study explores how seasoned speakers employ specific phonetic and visual strategies around—and during—turn-internal pauses for maximum effect. In a highly persuasive speech, Barack Obama, the former president of the United States and a charismatic presenter, demonstrated systematic use of two distinct pause type categories, out of which the most prominent were *post-message* pauses that featured (i) extreme pause durations that were paired with reflective facial gestures, (ii) steep pitch and intensity fall slopes right before the pause, and (iii) complete lack of phrase-final lengthening in the pre-pause speech material, which may be perceived by the listener as an increased speech rate before the pause. We speculate that the visual means invite the audience to be part of the contemplative process, while the extreme prosodic contrasts and the resulting abruptness signal the speaker's intention to convey charisma and related traits like self-confidence.

Key words: Pausing, silent pause, acoustic analysis, public speaking, persuasiveness, charisma, prosody, leadership.

1. Introduction

1.1. Phonetic characteristics of silent pauses

Politicians and professional public speakers know that skillfully used silent pauses not only mark prosodic boundaries. For example, in view of her results on politicians' speeches the French phonetician Danielle Duez writes that pauses "have an illocutionary function and serve to increase the strength of the argumentation" [1:643]. Moreover, they allow speakers to "maintain a balance between what was being said and what was left unsaid" [1:651]. Or, as phrased by presentation trainers like Huckle [2], a skillful use of silent pauses signals confidence, gravitas, and draws attention to one's speech. Unplanned, filled pauses or disfluencies may have a detrimental effect on one's oral delivery if their frequency of occurrence exceeds a certain threshold that depends on the filler's phonetic properties, see [3]. Planned pauses, by contrast—which are silent and purposeful—are often used for rhetorical and stylistic purposes.

In speech production studies, the use of planned pauses in public speaking traditionally has been viewed and researched mostly with regard to the properties of the pauses themselves, such as, for example, their total duration and frequency of occurrence. In acoustic studies of political speech and speakers, Duez already observed in 1982 that the more uninterrupted, monologue-like and pre-planned the speaking context is, the greater the duration of pauses [4]. Public speeches can therefore be expected to feature relatively long pauses, and news readings to have on average shorter pauses than poetry recitations, which is indeed in accord with recent evidence provided by [5]. Fifteen years later, Duez added to her findings that, with increasing authority and social standing, the pauses speakers make in their speeches become longer and more frequent [1]. The former French president Francois Mitterrand, for example, used shorter and fewer pauses in his speeches as a challenger and opponent, i.e., before he came to power, than in his later presidential speeches. More recently, Sturm & Volin [5] showed that pause durations can also vary depending on the speaker's tempo, particularly for uninterrupted, monologue-like and pre-planned speeches. The faster speakers talk in such speeches, the fewer pauses they make – and there is an additional tendency for the overall number of pauses to decrease.

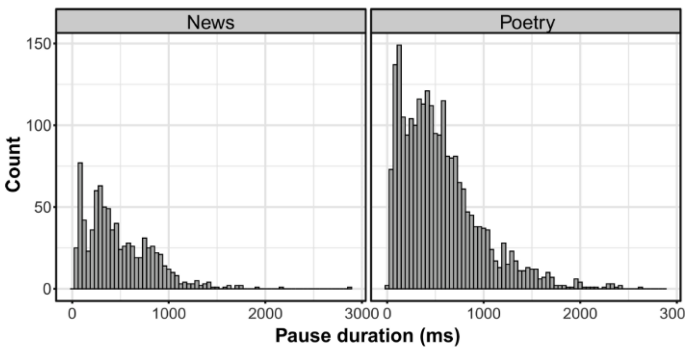

Figure 1: *Distributions of pause durations in news reading and poetry recitation (Czech speakers). Figure adopted based on a CC-BY 4.0 license from the paper of Sturm & Volin [5].*

In general, pause durations in speech have been shown to vary from very short (below 200 ms) to exceptionally long (above 2,000 ms), depending on speech type and modality. Campione & Véronis [6] in their corpus study on read and spontaneous speech of various speech genres have proposed a tripartite pause classification system with brief pauses (< 200 ms), medium-length pauses (200 ms to 1,000 ms), and long pauses (> 1,000 ms). In line with that, Sturm & Volin showed that distributions of pause durations across "speech genres seem to be bimodal, with one mode around 100–150 ms and another around 500–600 ms" [5:8]. An important implication of this statement is that, although there are pauses > 1,000 ms, they are extremely rare and hardly ever form a pronounced separate peak in a distributional analysis of pause durations, even for genres like news reading and poetry recitation. Figure 1 from Volin & Sturm [5] illustrates that.

In political speeches with planned rhetorical elements, however, even pause lengths of 2,000ms and above are common [1,7]. It is mainly these exaggerated "dramatic pauses" or "effect pauses" that rhetorical trainers stress as being particularly charisma-enhancing. Rehn writes, for example, that "'the pause for effect', the longer, measured pause that a great speaker can deploy to increase tension, underline a point, enhance a joke, and so on […] isn't an easy art, but it can be what separates an OK speech from a great one" [8]. Similarly, Esenwein and Carnegie state in their 2021 book The Art of Public Speaking: "Pause, and the lens gathers the heat. Your thoughts will not set fire to the minds of your hearers unless you pause" [9]. Soorjoo's presentation guide goes one step further and recommends a specific pause duration to its readers, while repeatedly reminding them that "Pausing is one of the most powerful and dramatic vocal techniques you can use during your pitch […]. A pause of 2 to 3 seconds may seem long to you, but your audience will barely notice it" [10:15, 115].

1.2. Researching the pause-charisma link

What the effect of pauses is on listener perception, and more specifically, what it means for the image of the speaker and whether the use of pauses is associated with charisma and persuasiveness has rarely been addressed in phonetic studies so far; and when it was addressed, it was mostly in a broader context, for instance, cross-culturally. Findings reveal substantial cultural differences in how pauses are linked to ratings of charisma or related traits, for example, by Palestinian and American listeners [11] or by French and Italian listeners [12], suggesting that speakers of different languages and cultures may employ and respond to a unique set of acoustic-phonetic cues when rating a speaker's charismatic impact. It fits with this initial evidence that Silber-Varod et al. [13] recently observed a pausing behavior in political interviews that was very different across languages and reminiscent of the traditional differentiation between "word cultures" and "silence cultures", see [14]. The results of [15] according to which L2 German speakers of Chinese origin rated speeches with a higher number of (longer) pauses as more polite, whereas the opposite applied to L1 German speakers points in the same direction.

As insightful as such cross-linguistic comparisons are, they largely ignore the variations in pause behavior and pause durations within a language – as well as the effects of such variations on how well listeners can follow the speech and rate the speaker. That is, their information value for applied rhetorical training is relatively low. Moreover, listener ratings are often elicited based on speech material that is a mix of spontaneous (public debates, interviews) and planned speeches, making it hard to draw reliable conclusions on how a phonetic feature (setting), including that of pauses, is linked to charisma and persuasiveness.

Our line of research will explore the use of pauses by primarily English speakers in planned public speaking contexts and, therefore, with a focus on planned pauses. The case study of Barack Obama's pauses that we present here marks the initial step into this new line research. As we describe below, one decisive novelty of this line of research is that we will look significantly beyond the pauses themselves (their counts and durations) and do perceptual and acoustic analysis of the multiparametric and multimodal phonetic characteristics of the pre-pausal and post-pausal speech sections in which the pauses are embedded. Investigating the role of pauses in public speaking is part of a larger research effort to pinpoint the specific acoustic features that seasoned speakers use to woo the audience—smart phonetic strategies that can ultimately make the audience change their attitude and behavior and that can potentially build any speaker's charisma and credibility if applied effectively , see Michalsky & Niebuhr [16] for an overview of the phonetics of charismatic vocal performances. Brillman & Scherer [17] demonstrated in detail how influential prosodic features are in winning the votes of an audience and that also machine-learning algorithms, when asked to predict the results of a voting, achieve a higher accuracy based on prosodic features than on, for example, text and video features.

Our definition of charisma follows that of Antonakis et al. who have characterized charisma as "values-based, emotion-laden leader signaling" [18:304]. There are small but crucial differences to other definitions of charisma, such as "the ability to attract and retain followers without the benefit of formal authority", put forward by Rosenberg and Hirschberg [19:640]. While their definition is without a doubt true, we agree with Antonakis et al. that, in order to avoid circularity, charisma should not be defined in terms of its effects. Moreover, like other earlier definitions, the one of Rosenberg & Hirschberg also ignores a recent key insight of charisma research, namely that charisma is not something a person *has*, but something a person *does*. It is a complex system of multimodal signals stemming from words, prosodies, gestures, facial expressions but also attire, age, body size, dialect, skin color, social status etc. Speakers convey these signals partly consciously and partly unconsciously; some types of signals can be controlled, while others cannot.

Determining how phonetic signals create perceived speaker charisma is by no means a simple task, yet it has received substantial research interest in the last decades as the practical need for it is growing—the sheer amount of content that is generated nowadays means that any technique that can make a speaker stand out and yield tangible results, i.e., change listeners' attitudes and decisions, is well worth the effort. For example, podcasts are on their way to "to becoming a true mass medium" [20] with average annual growth rates of more than 20 % in terms of both the number of podcasts and the number of listeners across many countries. Video posts on LinkedIn grow at a similar dynamic rate with the percentage of CEOs being active on such social media platforms alone having doubled since 2010, cf. [21]. Since 2021, associations such as the German Media Trainer

Foundation have started awarding an annual prize for the best CEO social-media video in Germany [22].

1.3. The phonetics of charisma

Specifically for English, which is the focus of this investigation, acoustic and perceptual studies of temporal and spectral aspects of charismatic delivery show that louder rather than quieter speech is associated with charisma [19]; or better yet, it is loudness[1] variability that recipients associate with charisma [23,24] or being a "good speaker" [25] and that make the most charismatic public speakers like Steve Jobs, the founder of Apple, stand out [26], while others, like Elon Musk, call themselves "a bad public speaker" [27].

With regard to how mean pitch levels of speakers are related to charisma and persuasiveness, findings are mixed—some studies show that lower-pitched voices elicit higher confidence and charisma ratings than higher-pitched ones [25,28,29] and that listeners prefer their leaders, both male and female, to have lower voices [30,31]. Other studies demonstrate that it is in fact the higher-pitched voices that are perceived as more charismatic [19] and are thereby, for example, more likely to win investor-pitch contests [32].

Often part of the explanation for such seemingly contradictory results is that the researchers and/or the raters had a different understanding of charisma. For example, some may think of charisma in terms of dominance and power. However, this idea of charisma is fundamentally different from the idea that relates charisma to charm, competence and emotions and that underlies the definitions of Rosenberg & Hirschberg [19] and Antonakis et al. [18] – as well as our own study, cf. 1.2. A further reason why charisma is likely claimed to be associated with both low and high pitch is that the temporal patterns of phonetic charisma parameters as well as their interaction in creating perceived charisma are insufficiently understood. That is, being a charismatic speaker could mean to use both high and low pitch levels but at different points within utterances and during a speech. For instance, Jiang and Pell [33] looked into the use of intonation patterns within utterances and found that confident speakers use a higher pitch in sentence-initial positions and a low falling intonation at sentence endings, cf. the concept of the audible "full stop" in Michalsky & Niebuhr [16]. Also note that the impression of being high or low is not solely conveyed through pitch but also through certain timbre characteristics and other prosodic features. In fact, we will come back to this point in the discussion of our results, where we propose the possibility that long, silent pauses may be a substitute feature for low pitch.

Results for tempo are similarly heterogeneous as for pitch. Fast speakers have been rated as more competent, knowledgeable [34] and charismatic [19], while slower speech rates are associated with character and composure [35], all of which are the many facets and building blocks of charisma. Some studies have shown that in fact a combination of both is best, and variations in tempo are the key to "good speaker" ratings [25,36]. We will also present the perspective based on our data that both fast and slow can be related to charisma in the sense of slowing down before one type of pauses and, in particular, not slowing down (even speeding up perceptually) before another type of pauses.

Where research does converge is the finding that pitch variability, or using a wide pitch range, is a major trademark of charismatic speech [19,26,37,38]. That is, the variability and versatility principle, described earlier for loudness, holds also for pitch.

Besides the features listed above, there are other, more nuanced and refined prominence elements that have been addressed by merely a few studies but that add substantially to speech versatility and, thus, to speaker charisma if used sparingly and skillfully. One such element is the emphatic accent chain, where every word in a sequence of words is amplified [26,39]; similarly, reinforcement is a type of emphasis whereby the onset consonant of a stressed syllable gets lengthened, often doubling in duration [39,40, see also Erickson and Niebuhr, this volume, 41] – a strategy that was found both in everyday speech for persuasiveness purposes and in experienced, charismatic speakers' delivery [42,43]. An exceptional example of the latter is Steve Jobs—an undoubtedly charismatic and persuasive presenter who utilized a wide variety of phonetic features and whose speech has been studied extensively, see [26].

Another world-class speaker is Barack Obama, the former president of the United States. Such speakers raise the bar for all presenters; they serve as role models for others and are a goldmine for researchers who aim to explore and understand persuasiveness-boosting phonetic features.

1.4. Aims and questions

When it comes to exploring and understanding persuasiveness-boosting phonetic features, Barack Obama's idiosyncratic approach to pauses is a case in point—not only does he expertly use pause length, but he also makes the speech material around the pause work hand in hand with the pause which "gives Barack Obama incredible poise, control, gravitas and authority as a presenter" according to Huckle [2]. Besides Huckle's description from the perspective of a rhetorical trainer, Obama's pauses have been topicalized in several scientific publications that range from co-creating pragmatic meanings with the audience [44] to predicting audience-response patterns in a machine-learning framework [45].

In fact, the systematic use of (long) pauses and impressive acoustic contrasts in the speech material that immediately precedes/follows the pauses are arguably one of the most prominent features of the former president's speaking style. However, to the best of our knowledge, despite their perceptual salience and assumed charismatic contribution and the attention that they received in previous studies, there is, as yet, no in-depth audio-visual analysis of Obama's pauses and their prosodic embedding. Addressing that gap in an explorative approach and based on a sample of pilot data is the aim of the present study. We base our study on one of Obama's recent major public speeches that has gained popularity and where the former US president aspired to be highly persuasive: It is the speech held at the 26th United Nations Climate Change Conference (COP#26) in Glasgow on November 8, 2021.

[1] Note that we want to make this article easily accessible across disciplines and therefore avoid untransparent technical terms such as fundamental frequency (f0) and short-term acoustic energy (intensity) – the main correlates of pitch and loudness – unless the context requires it.

We focus on the prominence of and relationships between (i) pause length, (ii) visual (i.e. mimic) cues during the pause, and (iii) prosodic patterns and dynamics in the pre- and post-pausal speech sections. The research questions that we posed were:

(1) *How does Barack Obama embed pauses phonetically and realize them visually?*

(2) *Do the phonetic/visual means vary with pause location/type of pause?*

2. Data and Methods

2.1. Speech sample

The speech that we analyzed, Barack Obama's address at the 2021 United Nations Climate Change Conference on the urgency of climate action was one of the most important and popular speeches held by the former president in recent years—for example, in terms of online views and comments. We used the video recording of the speech that is freely available online[2], see Figure 2. The audio signal of the speech was extracted and stored locally as an uncompressed WAV file.

Figure 2: *Screenshot of Barack Obama's UNCC speech at COP#26, Glasgow, Scotland/UK in 2021.*

The seriousness and urgency of the topic that Obama is personally very concerned and vocal about made his selected COP#26 address a fine example of a highly persuasive and fervent speech. Moreover, unlike his many other speeches, including pre-election and democratic-convention speeches or State-of-the-Union addresses, Obama's climate speech featured practically no applause or interventions from the audience. This was a fourth important criterion for us to select his COP#26 speech, in addition to the above-mentioned criteria of importance, popularity and the underlying intention to persuade. Pauses often serve to accommodate applause (cf. [44,45]), and this would have interfered in the present study with their analysis and interpretation as planned stylistic elements with communicative functions, cf. the works of [46] and [47] on pause functions in syntactic and discourse structure, or, more recently, of [48] and [49] on the role of pauses for cognitive states attributed to speakers in question-answer sequences.

Obama's COP#26 speech was almost 50 minutes long. For our analysis, we picked an about 20-minute long segment from the middle of the speech, i.e., after the initial exchanges of niceties and direct addresses had been given, and before any concluding interactions with the audience and entertaining bits had begun that do not directly relate to the message. Note that we have made the exact time stamps of the analyzed audio excerpt as well as our corresponding annotation and TextGrid files freely available upon request via Zenodo[3].

2.2. Analyses

2.2.1. Classification of pauses

First, we identified all silent pauses in Obama's 20-minute speech excerpt, combining an automatic-detection approach based on the script of De Jong and Wempe [50] with manual, auditory checks and corrections, for example, with respect to the automatically labelled boundaries of pause onsets and offsets. The minimum pause-detection threshold was set to 300 ms. This time interval can be considered standard in research on pauses. It is long enough for pauses to be reliably detectable by listeners and for voiceless-stop closures to get misclassified as pauses, cf. [51]; and, simultaneously, it is short enough to sufficiently cover pauses that listeners perceive as "fluent"— see [52] for an overview. Regarding the latter, as our focus was on planned pauses, our manual auditory checks of the automatic output were also made to exclude all pauses related to disfluencies and hesitations from our sample, including silent pauses adjacent to hesitation markers. We used PRAAT [53] and its TextGrid files to carry out this initial step.

All remaining pauses were, in a following step, labelled as instances of either "pre-message", "mid-message" or "post-message" pauses, based on semantic, syntactic and auditory criteria. Since we found only a few pre-message pauses in Obama's speech, those were excluded from the analysis, which left us with two pause types – mid-message (MM) and post-message (PM) – each of which was sufficiently frequently represented on Obama's speech. The sample size of MM pauses was N=46, and that of PM pauses N=36. So, overall the sample size was N=82. A speech excerpt that exemplifies these two types of pauses is provided in the following:

- *"Collectively and individually ___ (MM) we are still falling short ___ (PM). We have not done nearly enough to address this crisis ___ (PM). We are going to have to do more ___ (PM). And whether that happens or not to a large degree ___ (MM) is gonna depend on you ___ (PM)."*

All MM and PM classifications were conducted by the first author EB. MM pauses were, in terms of their classification criteria, characterized by following semantically and/or grammatically incomplete utterances and/or statements that gave, on a general auditory basis, the impression of being continued after the pause. By contrast, pauses were classified as being of the PM type when they separated cohesion- and coherence-wise complete and well-formed utterances and/or when they followed utterances that, on a general auditory level, gave the impression of not being seamlessly continued after the pause.

Thus, the semantic and grammatical criteria took into account the structures that immediately framed the pause on

[2] https://www.youtube.com/watch?v=kpOSlo_gwWo (last accessed 18 March 2023)

[3] https://magentacloud.de/s/A6Sn9ATYbSA5Ndf

both sides, whereas the auditory MM and PM classifications were primarily guided by the entire utterance before the pause. More specifically, EB relied on her holistic impression and the resulting expectation as to whether or not the utterance before the pause would be in a "good continuation" with the utterance after the pause, for example, in terms of grammatical structures or phonetic shapes. Note with respect to the internal validity of these auditory impressions that EB holds a M.A. degree in Teaching English as a Second Language from the University of Illinois at Urbana-Champaign (UIUC), USA, and a Ph.D. in second language speech acquisition from Bowling Green State University (BGSU), USA. Prior to coming to SSE Riga, she taught academic English, conversational English and American English pronunciation to international students at UIUC, BGSU, the University of Texas, and the Ohio State University.

In those cases in which semantic/syntactic classification criteria were ambiguous or clashed with the auditory criteria, we gave priority to the auditory criteria (given that it is the prosodic phonetics that we were interested in); and if the auditory criteria were also inconclusive as to the MM vs. PM classification, we excluded the ambiguous pause item entirely from the sample. Two examples of problematic cases are given below:

- *(1) And that's why I'm here today. ___ (PM). To talk about what's happened in the six years since I spoke to the United Nations.*
- *(2) … and to talk about the steps we need to take if we wanna keep doing big things. _____ (PM). Because when it comes to climate, time really is running out.*

From a semantic standpoint, we see the same ideas being continued across the pauses in (1) and (2). Moreover, syntactically, both (1) and (2) represent sequences of complete and incomplete sentences; for example, in (2), the conjunction "because" connects an independent clause with a dependent one. While, based on these characteristics, the pauses in (1) and (2) would belong to the MM category, the auditory impression on both examples was that of a complete and categorically-sounding initial statement to which, after the pause, a further explanation or justification had been added. Therefore, the examples (1) and (2) were finally classified as containing a "post-message" (PM) pause. In similar situations, often involving conjunctions like "and" or "but", we relied on the auditory impression to disambiguate the linguistic structure and determine the type of pause.

2.2.2. Acoustic analysis

The acoustic analysis was carried out in a semi-automatic fashion by the first author (EB). The following explanations describe the process chronologically.

First, we determined the pause boundaries, i.e., we labelled pause onsets and pause offsets in PRAAT, starting from the TextGrids that were automatically created by the pause-detection script of De Jong and Wempe [50], see 2.2.1. The script's output was refined such that onsets and offsets of pauses were set where the signal's intensity (RMS, dB) first fell by 10 dB or increased again by 10 dB relative to the mean intensity levels of the pre- and post-pausal sections, respectively (see [51] for these 10 dB thresholds). Note that the decreases/increases of 10 dB were measured with PRAAT's default analysis setting. That is, the pitch floor was 75 Hz, which corresponded to an analysis window of 40 ms over which the signal's elongations were integrated into the RMS calculation. Pause durations were measured as time intervals in between onset and offset labels.

The pre-pausal and post-pausal sections themselves were represented by the three syllables immediately before and after the pause—including unstressed syllables. The reason to go for syllable triplets on either side of the pause was the conclusion by Byrd et al. according to which, for English, prosodic "boundary effects extend from one to three syllables […] near the phrase boundary in both directions" [54:1589]. Further segment boundaries within these pre-pausal and post-pausal sections were placed following the recommendations for phonetic segmentation in Machač & Skarnitzl [55].

Next, the mean speaking rate in syllables per second (syll/s) was determined separately for the (i) pre-pausal MM, (ii) post-pausal MM, (iii) pre-pausal PM and (iv) post-pausal PM syllable triplets. To that end, we divided all time intervals (i)-(iv) by three. Figure 3 shows a combined sound-TextGrid example of an annotated MM pause (0.85 s) with its surrounding trisyllabic pre-pausal ("a success", 0.77 s) and post-pausal ("for the first" 0.50 s) speech sections.

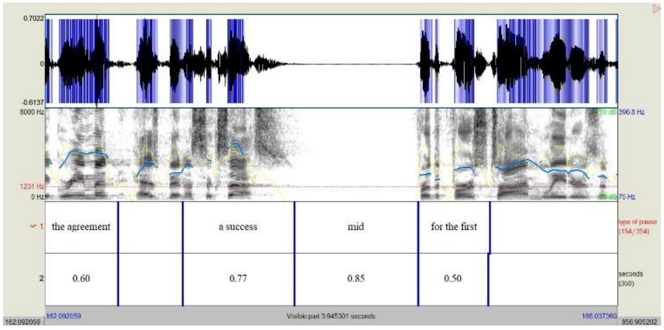

Figure 3: *Example of a combination of sound (waveform and spectrogram 0-8 kHz, top) and TextGrid files (text and duration tiers, bottom) for an annotated MM pause (referred to as "mid" in TextGrid tier #1) in Obama's analyzed speech.*

In addition, to measure the presentation's overall speaking rate , we randomly chose a sample of speech sections (i.e. interpausal units) from outside the pre-pausal and post-pausal areas in Obama's selected speech excerpt. With N=38 we took care that the size of this reference sample was similar to the MM and PM samples. The pause onset and offset boundaries in this reference sample were manually checked and corrected as in the main (test) sample. Then we determined the syllable-per-second rate for all of these 38 randomly selected speech sections.

Pitch measurements were done manually in Hz using PRAAT's autocorrelation algorithm (pitch floor = 75 Hz; pitch ceiling = 500 Hz). Three pitch features were measured within the syllable triplets before and after the pause: pitch range, the slope of the falling/rising pitch movement, and the pitch change across the pause. As is schematically illustrated in Figure 4, pitch range was the difference between the f0 maximum and minimum in the respective pre- and post-pausal sections. Pitch slope was the pitch range divided by the time it took to get from f0 maximum to f0 minimum (or vice versa, Hz/s); and the pitch change across the pause was the difference between the last f0 value before the pause and the first f0 value after the pause. The pitch changes were measured to determine to what degree Obama would resume his pre-pausal pitch level after the pause or strive for creating a pitch contrast or "discontinuity" from before to after the pause.

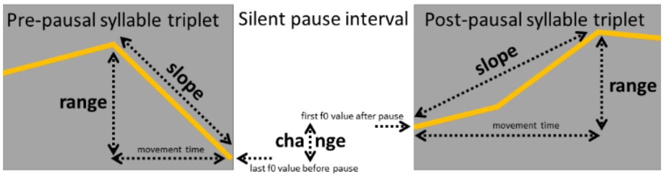

Figure 4: *Schematic illustration of range, slope, and change measures.*

F0 minima and maxima measurements were only taken inside vowel segments. If f0 peaks/valleys not were well-defined inside the vowel segments, for example, if f0 ran flat across the relevant vowel, then the f0 minimum/maximum was measured at the point closest to the slope in the direction of pause boundary; if two identically high or low f0 peaks/valleys occurred within a three-syllable section, then the one closest to the pause boundary was used, as it was assumed to have the most capacity to shape the auditory impression of the pre-pausal or post-pausal area.

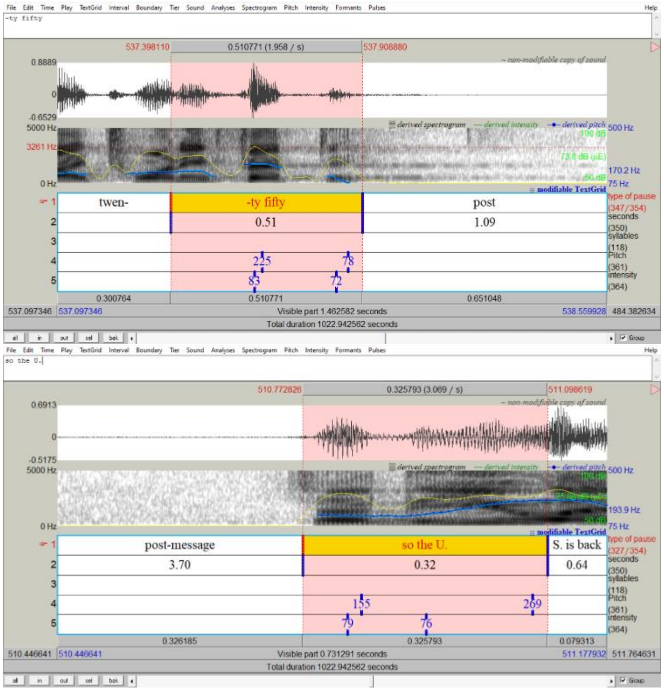

Figure 5: *Examples of manually determined pitch values (TextGrid tier #4) and intensity values (TextGrid tier #5) in pre-pausal (top) and post-pausal (bottom) PM contexts.*

For intensity, we measured the same set of three parameters in the pre-pausal and post-pausal syllable triplets as for pitch (see Figure 4): intensity range in terms of the RMS (dB) distance between the highest and lowest intensity level in the respective trisyllable; intensity slope in terms of the intensity range divided by the time interval in between the two extreme intensity levels (dB/s); and the intensity change across the pause, i.e. the difference between the last dB value before the pause and the first dB value after the pause.

Like for pitch, all intensity measurements were conducted manually in the syllable nuclei, using PRAAT's intensity tracker. Also like for pitch, we gave intensity values closer to the pause priority in the case of two or more identical values within the same pre- or post-pausal syllable triplet. Unlike for pitch, however, intensity ranges and slopes were only measured at peaks. Thus, the lowest intensity value corresponded to the lowest intensity peak in the respective trisyllable, not to an intensity valley or the end of an intensity fall. Figure 5 shows examples of sound-TextGrid combinations with manually determined Hz and dB measurements taken before (top) and after (bottom) PM pauses.

2.2.3. Visual Analysis

Confident, relaxed and expressive visual means in Obama's public addresses are common, and often, they appear to go hand in hand with his most dramatic, reflective pauses. We set out to explore which phonetic features these visuals are synchronized with, and, most importantly, whether their use is associated with a specific pause category, assuming that most visual prominence is reserved for the PM category.

To this end, all 82 MM and PM pauses in the 20-minute excerpt of Obama's COP#26 speech were analyzed in terms of facial expressions. Only the time intervals within the silent gap of the pause itself were taken into account, i.e., the pre- and post-pausal trisyllable sections were not included in the analysis. Future analyses will include a parameterization of facial expressions in terms of Action Units (AUs) defined by the Facial Action Coding System (FACS), see [56]. In the initial step of this current study, we limited the analysis to a binary division of pause-internal facial expressions into either "present" or "absent".

Like the acoustic analysis, the visual analysis of pause-internal facial expressions being "present" or "absent" was done manually by the first author EB. The video was paused at the onset of a pause, and then EB went through the pause in a frame-by-frame fashion (the analyzed COP#26 video of Obama features a 30 fps rate) and decided, based on her observations, whether there was a significant and meaningful change in facial expression during the pause. If such a change was found, she labelled visuals as being "present" for that pause. In all other cases, she labelled visuals as being "absent". Note that anticipatory speech gestures, such as lip rounding or jaw lowering (cf. [57]) were not considered meaningful changes in facial expression and, thus, did not trigger a "present" label.

Obama's facial gestures ranged from excessive blinking, lower lip curls, lip smacks, bites and tongue twirls behind cheeks to a rich combination of eye and lip gestures, see the provided examples in Figure 6. All of these features appeared to be used systematically and indicative of an internal, real-time thought process, seemingly inviting the audience to process what was said. Thus, it seems likely that they serve a stylistic, rhetorical function rather than a physiological need to ease the discomfort of a long pause.

Figure 6: *Examples of Obama's inter-pausal facial expressions.*

2.2.4. Inferential statistics

We ran a one-way MANOVA with Pause Type as fixed factor and all acoustic and visual measures as dependent variables. In addition, we conduced correlation analyses (based on Pearson's Product Moment Correlation Coefficients, PMCC) and linear discriminant analyses (LDAs) to shed light on the inner mechanics of Obama's MM and PM pauses as well as on their distinctiveness. The results of the all analyses are summarized below.

3. Results

3.1. Pause duration

Obama produces a very clear and also statistically significant difference between the durations of his MM and PM pauses ($F[1,80]=50.821$, $p<0.001$, $\eta_p^2=0.388$). More specifically, MM and PM pause durations are separated by about a factor-2 time difference. While it takes Obama on average 2,195 ms to make a PM pause, his MM pauses are on average only 1,013 ms long. However, the violin plot in Figure 7 also shows that the difference, although relatively large compared to pause categories in other studies (cf. [5,6]), is by no means categorical in nature. This is especially due to the MM pauses whose durational variability is about twice as big (s = 924 ms) as that of the PM pauses (s= 567 ms). This difference in variation magnitude is also significant according to a supplementary F test ($F[45,35]= 2.656$, $p=0.004$).

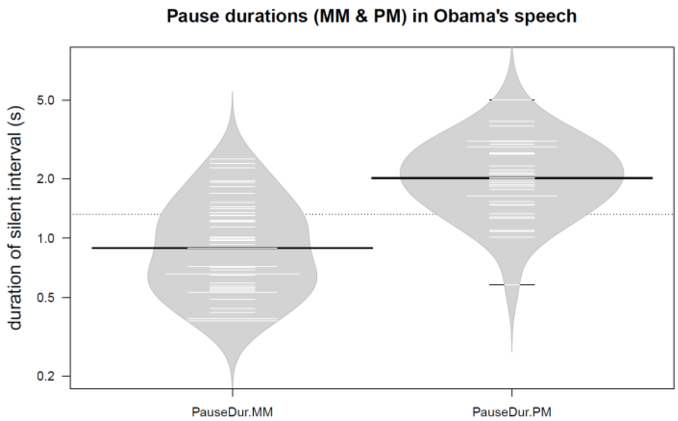

Figure 7: *Bean plots of the logarithmically scaled (s) MM (left, N=46) and PM (right, N=36) pause durations in Obama's analyzed speech.*

Furthermore, Obama's pause durations – even those of the MM pauses – are overall remarkably long, as is illustrated by the histogram in Figure 8, compared to a histogram we created after the data of silent pauses in other speaker's read speech performances from the study of Campione & Véronis [6]; see also the overview provided by Heldner & Edlund [58] and the study of Sturn & Volin [5], who report similar pause-duration distributions (and mean values for silent intervals of about 480-530 ms) both for news readings and poetry recitations. In all above mentioned previous studies, pauses > 2,000ms are "exotic" insofar as they represent less than 5% of the respective data sample. By contrast, in Obama's data sample pauses > 2,000 ms cannot be considered "exotic" at all. Rather, they account for 10% of the MM and even slightly more than 80% of the PM dataset.

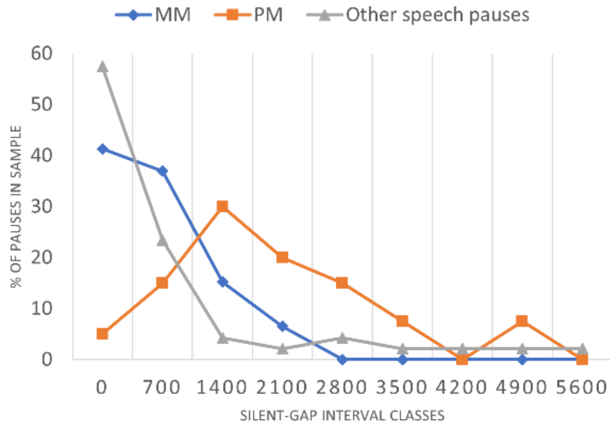

Figure 8: *Histogram of the relative (%) frequencies of silent-gap interval classes (width 700ms) for Obama's MM (blue) and PM (orange) pauses as well as for the read-speech pauses in Campione & Veronis [6].*

3.2. Temporal dynamics

The temporal dynamics is concerned with Obama's speaking rate in the three syllables before and after his MM or PM pauses and how this rate relates to Obama's overall speaking rate outside of these two pausal domains.

The central results regarding his temporal dynamics are summarized in Figure 9.

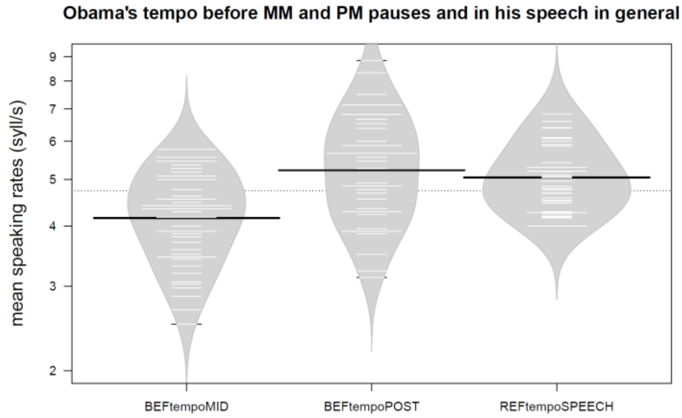

Figure 9: *Bean plots of Obama's speaking rates before MM and PM pauses as well as in his presentation outside pausal domains.*

Obama is significantly slower in the three syllables before his MM than before his PM pauses ($F[1,80]=20.282$, $p<0.001$, $\eta_p^2= 0.202$). When approaching a MM pause, Obama's rate is on average 4.25 syll/s, while it remains at about 5.40 syll/s before his PM pauses.

The previous wording already implies that Obama significantly slows down on the three syllables before MM pauses compared to an overall rate of about 5.12 syll/s in his speech outside pausal domains ($t[45,37]=-4.543$, $p<0.001$, $d=1.029$) – but he does not do so on the three syllables before PM pauses, where the speaking rate is statistically indistinguishable from the speaking rate outside pausal-domains. From a linguistic point of view, this means that, before his MM pauses, Obama's behavior is in line with the prosodic phenomenon of phrase-final lengthening, which is so robust and established that, according to Paschen et al. [59], it "is often assumed to be a universal process of languages", including American English [60]. Before his PM pauses, however, Obama remarkably violates this supposedly universal process of phrase-final, pre-pausal lengthening.

Note that Obama's overall speaking rate of slightly more than 5 syll/s matches well with the presentation tempo of other charismatic American-English speakers, such as Steve Jobs [26].

After the pause, the speaking rate that we measure on Obama's first three syllables is significantly higher than his overall rate outside the pausal domains. This applies in a statistically identical way to both the MM-pause (t[45,37]=2.512, p=0.012, d=1.029) and the PM-pause conditions (t[35,37]=3.740, p<0.001, d=1.029), see Figure 10.

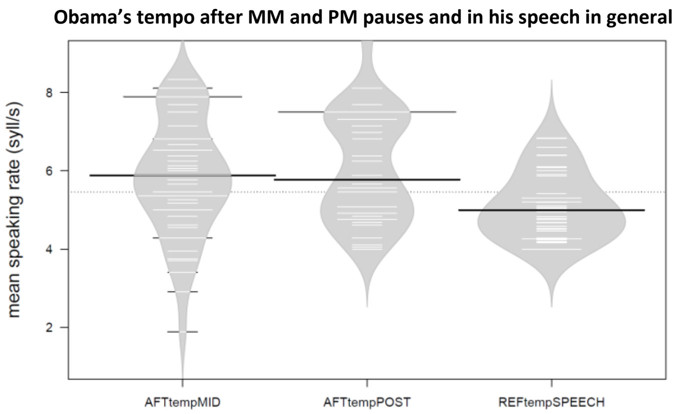

Figure 10: *Bean plots of Obama's speaking rates after MM (tempMID) pauses and after PM (tempPOST) pauses as well as in his presentation outside pausal domains (REFtempSPEECH).*

3.3. Pitch dynamics

Regarding the pitch dynamics, we observe a symmetry of the pre-pausal and post-pausal dynamic characteristics in Obama's speech, albeit with clear differences depending on whether these dynamics frame an MM or a PM pause.

One aspect of this symmetry is that both the pitch fall before the pause and the pitch rise after the pause have significantly smaller ranges for MM pauses than for PM pauses, see Table 1. For example, Obama's pitch falls on average 89 Hz before a PM pause but only by about 32 Hz before a MM pauses. This a difference of almost factor 3.

On the other side of the pause, i.e., for the post-pausal pitch rise, this discrepancy is smaller than factor 3, with the average rise being 80 Hz after PM pauses and 51 Hz after MM pauses. That is, the discrepancy in smaller mainly because the rise after MM pauses is larger than the fall before MM pauses. For PM pauses, by contrast, the falling and rising pitch movements are approximately equally large (89 Hz vs. 80 Hz).

Another aspect of symmetry is that the pre-pausal fall onsets and post-pausal rise onsets do not differ significantly in their relative timing between the two types of pauses. The pitch fall always sets in at a similar relative position within the three syllables, regardless of whether an MM or PM pause follows. The same applies to the rise onset after MM and PM pauses. Also, the durations of the rises and falls do not differ significantly depending on pause type. This applies both within the PM and MM rises and falls as well as between the rises and falls themselves. So, all pitch movements around the pauses always vary in a similarly short duration range between about 257 ms (fall duration before PM pauses) and about 292 ms (fall duration before MM pauses). Note that the shortest pitch-movement duration is that of the fall before PM pauses, which also shows the largest mean pitch range of 89 Hz.

Table 1: *Test statistics of the MANOVA (MM vs. PM) related to the pitch measures (BEF = before, AFT = after).*

Measure	F[1,80]	p	η_p^2
BEF fall range	20.702	<0.001	0.206
BEF fall time	1.128	0.291	0.014
BEF fall slope	15.556	<0.001	0.163
AFT rise range	13.219	<0.001	0.142
AFT rise time	0.130	0.720	0.002
AFT rise slope	8.116	<0.001	0.092
Pitch change across pause	0.219	0.641	0.003

In contrast, Obama's pitch dynamics are characterized by clear differences between the movement ranges in MM and PM contexts, while the overall durations and onset timings of these movements remain comparable. As a consequence, the third characteristic in Obama's pause-related pitch dynamics is that pitch slopes show considerable differences in MM and PM contexts. In particular, the slope of the pitch fall before PM pauses is on average 418.2 Hz/s and hence orders of magnitude steeper than the fall before MM pauses, whose pitch slope is on average just 155.6 Hz/s; see Figure 11. The effect size of this difference is one of the largest of the MANOVA (see Table 1). Moreover, within the value span of one standard deviation, the PM pitch falls are on average 350-500 Hz/s steep, which corresponds to values that other speakers reach in emphatic pitch-accent realizations [40,60].

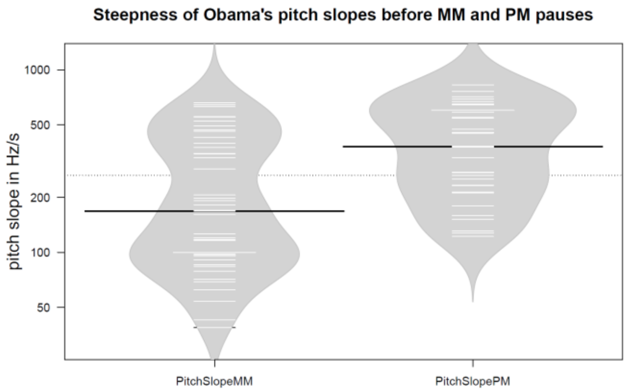

Figure 11: *Bean plots of the falling pitch slopes before MM pauses (right) and PM pauses (left).*

The rising slopes after the MM and PM pauses are also significantly different, but with the effect size being significantly smaller than for the falling slopes (Table 1). Again, however, the slope of the pitch rebound is considerably steeper after PM pauses (335.5 Hz/s on average) than after the MM pauses (213.1

Hz/s on average). Note that even the latter, rising MM slope of 213.1 Hz/s is still steeper than the falling slope before MM pauses, which is – with 155.6 Hz – the flattest of all.

We find no significant difference regarding the pitch continuity across MM and PM pauses, i.e. the pitch change from the last f0 value before the pause to the first f0 value after the pause is roughly equally pronounced for MM and PM pauses. In both cases, the pitch change manifests itself in a slight increase of on average 16 Hz (MM) or 11 Hz (PM) from before to after the pause. In fact, at least in view of the mean values, the pitch continuity is even more maintained across PM pauses than across the MM pauses; or, in other words. PM pauses are clearly not associated with bigger pitch breaks than MM pauses.

3.4. Intensity dynamics

The intensity dynamics matches the overall pattern of the pitch dynamics. That is, the drops and rebounds in intensity are symmetrical on both sides of the MM and PM pauses insofar as they are approximately similarly long and equally timed within their respective three syllables. In addition, the intensity ranges in the context of MM pauses are significantly smaller on both sides of the pause than in the context of the PM pauses, see Table 2. Similar to the pitch dynamics, the greatest intensity dynamics occurs before PM pauses.

Measure	F[1,80]	p	η_p^2
BEF drop range	20.352	<0.001	0.203
BEF drop time	0.002	0.961	0.000
BEF drop slope	10.195	0.002	0.113
AFT rebound range	6.744	0.011	0.078
AFT rebound time	2.274	0.136	0.028
AFT rebound slope	7.713	0.007	0.088
Intensity change across pause	8.836	0.004	0.099

Table 2: *Test statistics of the MANOVA (MM vs. PM) related to the intensity measures (BEF = before, AFT = after).*

In terms of concrete numbers, we measure intensity drops before pauses whose resulting mean value is almost twice as high (13.6 dB) in the PM context as in the MM context (7.6 dB); and that contrast gets still greater after pauses. With 6.1 dB, the intensity rebound in the PM context is almost three times as high as in the MM context, where it averages only 2.9 dB, see Figure 12. The latter MM intensity rebound is on average also slightly shorter (188 ms) than the other three MM/PM intensity rebounds, all of which last approximately 220 ms. But, like for the pitch movement durations and timings, these intensity differences are not statistically significant.

For the intensity slopes before and after pauses we find significant differences between the MM and PM conditions as well. As is shown in Table 2, this applies to both the pre-pausal and the post-pausal context. Note that, unlike for the pitch dynamics, the relative differences in intensity dynamics are greater after pauses than before pauses. After pauses, the average rising slopes are 16.3 dB/s in MM and 42.1 dB/s in PM contexts. This almost corresponds to a factor-3 difference. Compared to that, the average falling intensity slopes before pauses differ with 42.3 dB/s in MM and 73.2 dB/s in PM contexts by less than a factor of 2. In absolute values, however, the pre-pausal intensity dynamics is still clearly higher than the post-pausal dynamics – with the absolute highest dynamics characterizing the pre-pausal PM context, just as for the pitch dynamics.

Another similarity to the pitch dynamics concerns the size of the intensity break across pauses. The change in dB from the last intensity measurement before the pause to the first intensity measurement after the pause is on average larger in the MM context (9.6 dB) than in the PM context (4.4 dB). This is a significant difference, see Table 2.

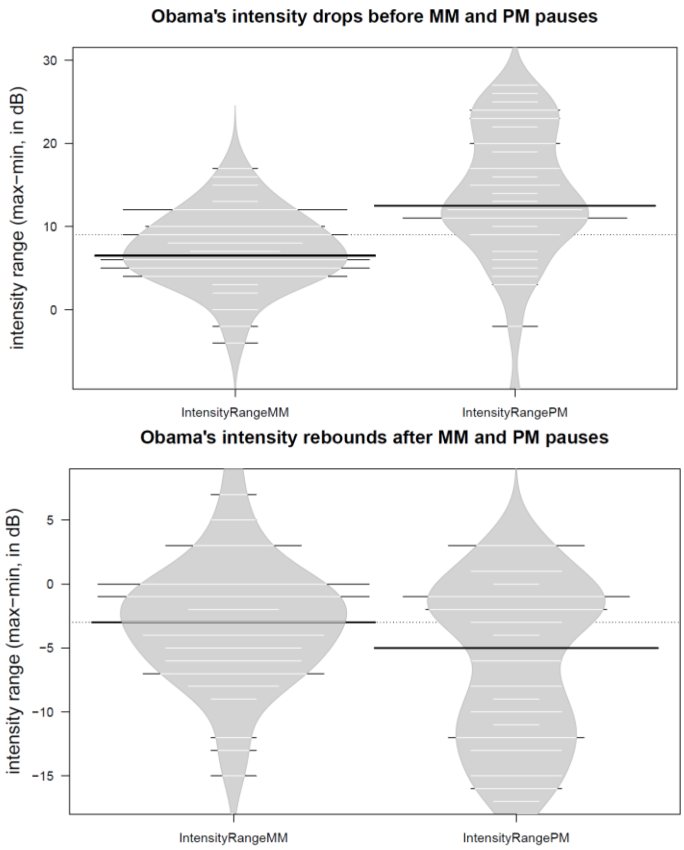

Figure 12: *Bean plots of the ranges of the intensity drops (top) and the intensity rebounds (bottom) measured MM pauses (right) and PM pauses (left) contexts.*

So, on the whole, if we combine pitch and intensity results, we see that the PM context is generally characterized by greater dynamics both before and after the pause while, at the same time, it maintains a greater continuity of both pitch and intensity across the pause. Further, more general results patterns are dealt with in terms of correlations in 3.5.

3.5. Overarching patterns: correlations

Finally, with regard to the internal mechanics of the pitch and intensity dynamics in Obama's speech, correlation analyses show that the slopes of pre-pausal pitch and intensity falls are positively correlated with their corresponding ranges in both contexts, MM (pitch: r[44]=0.873, p<0.001; int: r[44]=0.753, p<0.001) and PM (pitch: r[34]= 0.640, p<0.001; int: r[34]=0.703, p<0.001). That is, pitch slopes get steeper when pitch ranges get larger, and the same is true for intensity. In PM contexts, we additionally find the falling slopes before pauses to be negatively correlated with their fall durations (pitch: r[34]=-0.675, p<0.001; int: r[34]=-0.358, p<0.05). That is, when pitch falls get shorter, their slopes get steeper.

Taken together, these correlations underscore the central "goal" underlying the prosody before Obama's PM pauses: to create as high a dynamic as possible, i.e., to realize the most abrupt, steep pitch and intensity falls possible and then to contrast these abrupt changes with a particularly long silent gap. At least in terms of correlations, such an underlying "goal" seems not to exist in Obama's post-pausal prosodic patterns – and certainly not in the prosodic patterns on either side of Obama's MM pauses. In that respect, Obama's pre-pausal PM prosody is truly special.

The prosodic patterns on either side of MM pauses are, however, also associated with two significant correlations. Specifically, in MM contexts, the pitch change, i.e. the step from the last f0 value before the pause to the first f0 value after the pause, is negatively correlated with both the pre-pausal pitch range (r[44]= -0.437, p=0.002) and the pre-pausal pitch slope (r[44]=-0.357, p=0.015). That is, the higher pitched Obama starts speaking after a MM pause (compared to before the pause), the shorter and shallower he realizes the pitch fall before the pause.

That the pitch and intensity differences immediately before and after pauses are greater for MM than for PM pauses (cf. 3.3 and 3.4) does not rule out that the two negative correlations above are indicative of a higher-level mechanism geared towards creating a "good continuation" of prosody across MM pauses in a way that is absent for PM pauses.

3.6. Visual signals

As was stated in 2.2.4., we looked for facial communication gestures such as blinking, lower lip curls, lip smacks, bites or tongue twirls behind cheeks in Obama's pauses and, as a first basic measure in this initial analysis, classified all MM and PM pauses into two groups: those in which such facial communication gestures were present and those in which they were absent. The underlying assumption was that PM pauses are more often filled by facial communication gestures than MM pauses.

Results of this visual analysis show that this assumption is clearly met by the data. While only a minority of 17.4 % (i.e. 8 out of 46) of Obama's MM pauses are filled by facial communication gestures, the PM pauses contain such gestures in 72.5 % or in 26 out of all 36 cases analyzed here. This difference is significant ($F[1,80]=35.119$, $p<0.001$, $\eta_p^2=0.305$).

Figure 13 provides an overview of how the use of facial communication gestures is distributed over the durations of the MM and PM pauses. To that end, the MM and PM pause duration ranges were divided into their respective four quartiles (MM: 0-560 ms, 561-890 ms, 891-1,302 ms, 1,302-2,527 ms; PM 0-1,505 ms, 1,506-2,015 ms, 2,016-2,800 ms, 2,801-5,030 ms). We see that even within the MM pauses, none of Q1 is filled with facial communication gestures. In Q2 it is only 8.3 %. Specifically, there is only one MM pause lasting less than a second, in which facial communication gestures are found. Most MM pauses with facial communication gestures have durations in the fourth quartile. Typically, these pauses are about 2 seconds or longer. PM pauses show a similar pattern: most cases with facial communication gestures are in quartiles Q3 and Q4, demonstrating that also in this type of pause, the proportion of facial communication gestures increases significantly with a pause duration threshold of 2 seconds.

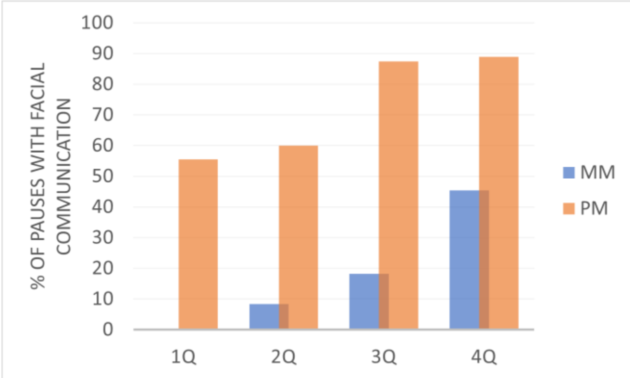

Figure 13: *Distribution of facial communication gestures across MM and PM pauses with different duration classes representing the respective quartile ranges 1 to 4.*

3.7. Distinctiveness of MM and PM Pauses

In order to estimate how clearly and systematically Obama distinguished the MM and PM pauses along the analyzed acoustic and visual parameters in his speech, we performed a linear discriminant analysis (LDA). The LDA included all those parameters for which the MANOVA showed significant differences between the MM and PM pause categories. In addition, we performed three variants of the main LDA, which were based on (a) only the measured temporal and intensity dynamics *before* the pause, (b) only the measured temporal and intensity dynamics *after* the pause, and (c) only the measurements *within* the pause (silent interval duration, intensity change from immediately before to after the pause, and pause-internal facial communication gestures).

Table 3: *Summary table of the main LDA's performance of MM and PM pause classifications both in absolute and relative terms.*

Main LDA: All significant acoustic and visual measures					
MM vs. PM			Predicted pause type		Total
			MM	PM	
Original pause type	Abs. frequ	MM	42	4	46
		PM	5	31	36
	%	MM	91.3	8.7	100.0
		PM	13.9	86.1	100.0

Table 3 shows that, with all 12 significant before-, within-, and after-pause measures included, the LDA is able to predict with an overall accuracy of 89.0 %, whether a pause in Obama's speech is designed for a MM context or for a PM context. That is a remarkable and statistically significant classification performance (Wilks-λ=0.408, $\chi^2[12]=66.257$, $p<0.001$) that stresses how systematically differently Obama realizes his MM and PM prosodic pause patterns. Also, note that the prediction accuracy is similarly high for both the MM pauses (91.3 %) and the PM pauses (86.1 %). In terms of the canonical discriminant coefficients, the top-5 contributors to this overall prediction accuracy come from the pause duration itself (0.444) as well as from the intensity change from before to after the pause (-0.460), the pitch range before the pause (0.365), the intensity slope after the pause (-0.276), and the pitch range after the pause (-0.240).

The results of the other three additional LDAs are summarized in Tables 4-6. Firstly, when only temporal and intensity dynamics before the MM and PM pauses are taken into account, i.e. 5 of the total 12 parameters, the LDA performance is lower, but still significant (Wilks-λ=0.566, $\chi^2[5]$=44.054, $p<0.001$) and close to 80 % accuracy in total (79.3 %), see Table 4. It is the ranges of pitch (0.641) and intensity (0.632) that contribute most to this accuracy, in combination with the speaking rate before the pause (0.591).

Table 4: *Summary table of the LDA's performance of MM and PM (absolute and relative) pause classifications when only pre-pausal parameters are taken into account.*

LDA only with acoustic measures from before the pause					
MM vs. PM			Predicted pause type: MM	Predicted pause type: PM	Total
Original pause type	Abs. frequ	MM	36	10	46
		PM	7	29	36
	%	MM	78.3	21.7	100.0
		PM	19.4	80.6	100.0

The LDA that was carried out based on the four significant post-pausal parameters performed worse than the one based on the five pre-pausal parameters, see Table 5. Yet, MM and PM prediction accuracies are still well above 50% chance level and amount to a total prediction performance of 73.2 % (Wilks-λ=0.821, $\chi^2[4]$=15.397, $p=0.004$). Note that it is primarily the MM pause type whose prediction accuracy decreased based on the post-pausal parameters. The most powerful parameters were the pitch range after the pause (0.796) and the intensity slope after the pause (0.563). That is, both pitch and intensity played a role after the pause – as well as before the pause.

Table 5: *Summary table of the LDA's performance of MM and PM (absolute and relative) pause classifications when only post-pausal parameters are taken into account.*

LDA only with acoustic measures from after the pause					
MM vs. PM			Predicted pause type: MM	Predicted pause type: PM	Total
Original pause type	Abs. frequ	MM	32	14	46
		PM	8	28	36
	%	MM	69.6	30.4	100.0
		PM	22.2	77.8	100.0

Finally, if the LDA is carried out only with those three significantly different MM vs PM parameters that occur within the pause and at the pause onset/offset, then the performance *increases* again to almost that level of accuracy that was achieved by all 12 significantly different MM vs PM parameters, see Table 6.

In specific, this means that overall prediction accuracy is significant and at 84.1 % (Wilks-λ= 0.551, $\chi^2[5]$=46.753, $p<0.001$). In fact, for the MM pauses, the accuracy is exactly the same as with the full set of 12 parameters. That the overall accuracy is about 5 % lower at all is only due to the PM pauses. They are about 11 % less well predictable only based on silent interval duration, intensity change from immediately before to after the pause, and pause-internal facial communication gestures. The first and the last parameter, i.e. two truly pause-internal ones, turned out to be most powerful in this LDA variant (canonical discriminant coefficients: silent gap duration = 0.678; facial gestures = 0.431).

Table 6: *Summary table of the LDA's performance of MM and PM (absolute and relative) pause classifications when only pause-internal parameters are taken into account.*

LDA only with pause-internal/adjacent parameters					
MM vs. PM			Predicted pause type: MM	Predicted pause type: PM	Total
Original pause type	Abs. frequ	MM	42	4	46
		PM	9	27	36
	%	MM	91.3	8.7	100.0
		PM	25.0	75.0	100.0

4. Discussion

4.1. Summary and interpretation

Several public-speaking trainers and guidebooks celebrate Barack Obama as the master of pauses in presentations. For example, while according to Huckle, a normal "presenter rattles through his/her content and slides like a runaway train", the special way of pausing in Obama's speeches gives him "incredible poise, control, gravitas and authority as a presenter; […] he uses pausing to seemingly simplify his message whilst at the same time magnifying the significance of its meaning" [2]. In the present study, we wanted to examine, substantiate, and objectify these and similar impressionistic descriptions through phonetic measurements.

The overall picture of our results indeed characterizes Obama's pauses as remarkable prosodic constructions; remarkable, because, for one thing, Obama's PM pauses in particular show phonetically uncommon patterns. With on average over 2 seconds – and in some cases well over 5 seconds – his PM pauses are longer than 95 % of other (silent) pauses that studies found in the read and spontaneous speech of "normal" speakers; see, for example, the measurement distributions in [5,6,58]. For another thing, Obama also does not slow down when approaching his long PM pauses. The speaking rate stays high even in the last 3 syllables before the pause. Lehiste [61] observed that English listeners expect such final (i.e. pre-boundary) lengthening of syllables; and according to Weirich and Simpson, the absence of this lengthening can cause a perceptual reversal of this expectation: "perception [of speech tempo] is strongly dependent on listeners' (linguistic) expectations. English listeners expect final lengthening. If this is not perceived, the stimulus is rated as shorter than the one before even if it has the same duration." [62:10]. This means that listeners hear Obama's speech rate as accelerated towards his PM pauses.

The resulting abruptness is further reinforced auditorily by the fact that Obama produces deep and steep falls in prosodic parameters before his PM pauses. This concerns both the voice pitch level and the intensity. At least for pitch, the dynamics of these falls is similar to those that occur with strong, expressive pitch accents on emphasized words. Together with the correlations that we found within the prosody of these falls (larger and shorter falls mean steeper slopes), it is reasonable to assume that the underlying objective before PM pauses is to maximize the signal dynamics. This finding is in line with Ohala [63], who found that a sharp terminal fall in pitch is what makes the listeners perceive the speaker as more powerful and self-assured.

The abruptness generated in tempo, pitch and intensity as well as to a certain extent also the discontinuity due to the long pause duration itself are, on the one hand, in symmetrical agreement with the similarly steep pitch and intensity rebounds as well as with a further increase in speaking rate after PM pauses. On the other hand, these dynamics and discontinuities are in contrast to our finding that pitch and intensity changes from before to after the pause are small for PM pauses: *That is, overall, PM pauses are acoustically characterized by combining a mirror-inverted high prosodic dynamics on both sides of the pause with a high prosodic continuity across the pause's long silent gap. The exact opposite applies to MM pauses.*

In the pre- and post-pausal sections of MM pauses, the pitch and intensity drops are shallower and less extensive than for PM pauses. A further notable difference to PM pauses is that the MM pauses' pre-pausal sections do show that final (pre-boundary) lengthening that is well documented for American English. Also, in terms of the magnitude of this phrase-final lengthening, Obama's tempo reduction (relative to the tempo outside pausal regions) is on average about -17 %, which is well within the range of reference values reported for "normal" speakers in previous studies that estimate the amount of phrase-final lengthening in American English to be between 7-18 %, cf. [64]. Correlations pointing to the underlying objective of creating abrupt prosodic changes in pre- or post-pausal sections also seem not to exist in the same way for MM pauses as for PM pauses. Instead, we found correlations whose underlying objective could be interpreted as creating a "good continuation" – to use term from Gestalt theory [65]: The higher-pitched Obama starts speaking after a MM pause compared to before the pause, the shorter and shallower are his pitch falls before the pause. Like for the PM pauses, but in the inverse way, this striving for a "good continuation" is in contrast to the pitch and intensity changes from before to after the pause, which emerge as strongly pronounced breaks for MM pauses (while PM pauses seem to strive for a "good continuation" in this prosodic aspect).

The high acoustic distinctiveness indicated in the above summaries of the MM and PM pause prosodies is supported by the results of the performed LDAs. Based on the measured acoustic parameters in the pre-pausal syllable triplets alone (i.e. without taking the actual pausal intervals into account), about 80% of the items in our sample could be correctly identified as MM or PM pauses. That is, it is already in his pre-pausal speech section that Obama prepares his audience for whether a PM or MM pause is going to follow. For the post-pausal syllable triplets, the identification accuracy was still at almost 75 %.

The MM and PM pauses differed significantly also with regard to the use of visual means, i.e. facial expressions: in the majority of Obama's PM pauses (about 75 %), the silent gap was filled by facial communication gestures. For the MM pauses, this only applied to a small minority of less than 20 %. However, whether or not facial communication gestures occurred was not evenly distributed within MM and PM pauses. Rather, for both pause types their occurrence was more frequent for longer pauses. This casts doubt on whether visually filled silent gaps are a defining feature of PM pauses. Rather, facial communication gestures inside pauses could be a general feature of Obama's speech that is more likely to be used the longer the pauses are produced. Irrespective of that open question, the LDA analyses showed that pauses – and MM pauses in particular – could well be identified solely based on pause duration and the use (or absence) of visual facial communication gestures. In the case of the MM pauses, these two features alone achieved a discrimination accuracy of more than 90 %.

4.2. Limitations and implications

What are the implications of the results of our study? Our results do not allow any direct conclusions for, for example, practical applications in public-speaking training. However, given that Obama is generally regarded as a very charismatic speaker, it is at least reasonable to assume that his special way of pausing – and realizing PM pauses in particular – does not run counter to this perceived speaker charisma. Rather the opposite seems to be true: as the quote from Huckle [2] and similar examples suggest, Obama's pausing behavior is often considered an integral part of his perceived speaker charisma.

Why should that be so? Long pauses could help the speaker maintain a healthy balance between high-pitched and low-pitched elements, both of which are essential for a persuasive delivery. Expert speakers know that a wide pitch range is key to a dynamic and powerful presentation: Steve Jobs featured an impressive pitch range, which was at least half an octave larger than that of the average speaker [26], and so does Barack Obama in this study. Besides the actual range, the distribution and the relative proportions of the high-pitched and low-pitched sections also matter. Both our current findings and studies like those of [25], [33] and [37] demonstrate that conveying persuasiveness and certainty is done by using higher pitch in the prominent initial position of an utterance (like a prosodic phrase) and all throughout the utterance, but not towards the utterance ending, where a low pitch is most beneficial for conveying persuasiveness and certainty.

Obama's utterance intonations follow exactly that recipe. He has a very high mean pitch level, which, statistically, is in the lower third of the range of pitch levels among female speakers of American English, both in read speech and in spontaneous speech, see [66,67]. In our analyzed speech sample, it was on average about 195 Hz (s = 30 Hz). This value is on a par with Steve Jobs' pitch level, which averaged just over 200 Hz in his keynote speeches [26]. While some of the observable pitch increase can be attributed to the specifics of giving speeches to massive audiences in spacious rooms, which requires greater phonetic effort and thus the increased loudness inevitably brings the pitch up, the fact that a bigger proportion of their speech is delivered at a higher rather than lower pitch remains.

At utterance endings, however, his voice pitch drops low, in particular before PM pauses. What role do the pauses themselves play in this? We put forward the assumption here that silent pauses resonate in the ears of listeners at that pitch level that preceded the pause; in other words, they serve as an "extension" of pre-pausal pitch. In that case, the extended silence would be a special strategy of Obama to give, in a charismatic speech's rich and contrasting intonation, the low pitches more perceptual weight – so that the low-ringing extended PM pauses can counterbalance the surrounding high-pitch sections and, on part of listeners, trigger exactly those impressions of "control, gravitas and authority" [2] that are required for a persuasive speech and that public-speaking trainers and guidebooks often recommend using low-pitched voices for, see, for example, Lang's book about "The Secrets of Charisma" [68:165].

Two arguments support this assumption: First, Obama's silent pause durations are shorter when the pitch level adjacent to the pause (especially before the pause) is less deep – as in the case of the MM pauses, for example. Secondly, we have found that the pitch continuity from before to after the pause is more strongly maintained across PM pauses than across MM pauses. We know from perceptual psychology that such a continuity is essential for the brain to fill gaps perceptually, for example, in the phi phenomenon or the phonemic-restoration effect, cf. [65,69]. That is, the f0 continuity across Obama's PM pause could be decisive for keeping the pre-pausal low pitch "ringing" in the listeners' ears during the pause. In contrast, the significantly larger f0 discontinuity on both sides of the MM pauses could prevent or weaken such a perceptual filling-in of the silent gap.

We assume furthermore that, in the case of Obama's PM pauses, the impression of "control, gravitas and authority" is reinforced also by the pre-pausal sections' sudden drop to a low pitch level and the post-pausal sections' low-pitched onsets. According to Ohala, confident voices exhibit sharp and dramatic pitch falls, whereby an initial higher pitch "is there in order to make the terminal fall seem to be even steeper, i.e. by virtue of having fallen from a greater height" [63:329]. Finally, even the abruptness in the form of the lack of phrase-final lengthening itself seems to be suitable for enhancing the impression of "control, gravitas and authority".

These assumptions are currently tested in a perception experiment with systematically varied stimuli from US politicians, in which we examine the effects of, in particular, the pause duration, the f0 continuity across the pause, and the ranges and slopes of pre-pausal pitch falls on perceived speaker charisma and related traits. We are confident that the results of these experiments can provide initial guidelines for an improved rhetorical practice.

Furthermore, our case study raised the question of how body language is coordinated with pauses and pause types and how such a coordination can result in different functions. Currently, research on this topic is sparse – and often focuses more on dialogue communication and arm or hand gestures than on facial expressions and monological performances such as presentations. Noteworthy is Beattie & Aboudan's remark that a substantial proportion of (e.g., hand/arm) gestures occurs inside pauses or begins during pauses; and if this proportion is put in relation to the total durations of pause and speech intervals in everyday language use, then "gestures originate more frequently in pauses than during actual articulation" [70:261] – especially if the speech material in question comes from social interaction. As regards the nature of the pause-gesture coordination, empirical evidence suggests that gestures (especially iconic ones) associated with pauses have a guiding function. They prepare content that will be communicated verbally after the pause [70]. It fits in with this idea that Esposito and Marinaro [71] found articulatory gaps during disfluent cognitive pauses to be highly (but age-dependently) positively correlated with gestural gaps, i.e. a freeze of all gestures. The occurrence of these cognitively driven pauses was so precisely predictable from a gesture freeze that Esposito and Marinaro speak of a gestures-and-speech partnership and conclude that "speech and gestures reflect a unified communicative planning process in the production of utterances" [71:13].

Projecting findings like these onto Obama's speech suggests that his use of facial communication gestures during pauses prevents the latter from sounding disfluent, the very long ones in particular; and that the overall high frequency of occurrence of facial communication gestures in Obama's pauses underscores his audience-oriented, i.e. socially driven way of speaking. We plan a perception experiment in which we compare audio-only stimuli with A/V and video-only stimuli from Obama's speech to further investigate the extent to which his facial communication gestures within silent pauses affect his perceived charisma and similar traits. The audio-only stimuli exclude the pause-internal facial expressions, whereas the video-only stimuli exclude the pauses themselves, at least in terms of their acoustic gaps. In a parallel branch of this perception experiment, comparable stimuli are generated with an unknown person (actor) in order to eliminate the context factor of the person "Obama" from the rating behavior of participants.

Finally, the fact that our study is a case study with a comparatively small sample bears a number of further implications. In follow-up studies, we first want to expand our sample so that we can carry out a more differentiated prosodic and visual analysis that, amongst other things, separately examines those cases with semantically, grammatically and/or auditory incongruent MM vs. PM classification criteria. In addition, within the MM and PM pause categories, we want to look for prosodic and visual differences depending on the surrounding morphosyntax. For example, it seems possible from inspections of individual cases that PM pauses and their adjacent pre- and post-pausal sections are (within the general patterns presented above) realized differently, depending on whether it is the utterance before or after the pause that is grammatically incomplete – or when both utterances are ellipses. In connection with this, we will also arrive at substantial statistics about how MM and PM pause categories co-occur with semantic and grammatical structures.

Future analyses could also be expanded to take other speeches by Obama into account, for example, speeches that are more entertaining or interactive in nature, such as Obama's annual White House Correspondents' Association Dinner speeches. Navaretta [45] conducted a machine-learning experiment based on Obama's pause durations and his concomitant hand, face, and head gestures in these dinner speeches, with each gesture type being roughly classified into 4-9 categories. As the results show, pause durations larger than 60 ms and their associated patterns of gesture labels were sufficient to significantly predict the audience's response linked to the pause in terms of six different categories: cheers, laughter, applause, cheers/applause, laughter/applause, cheers/laughter. That is, Obama's pauses seem to vary systematically and, moreover, such that they are designed to accommodate a specific, anticipated reaction. Similarly, Capone reports in view of an analysis of Obama's "South Carolina Speech" that "there is a perfect synchrony between the speaker and the audience, as the audience is able to distinguish short from long pauses and does not disturb or disrupt the main speaker's [i.e. Obama's] words." [44: 2968].

To also analyze such speeches with regard to measures of prosodic detail like those of our case study and to expand the visual ("body language") channel to include quantified mimic and gestural parameters beyond descriptive label systems such as FACS (e.g., via software apps like AFFDEX, cf.[72]) will bring further, more in-depth insights into how acoustic prosody and facial and gestural communication signals interact in form, function and time – insights that will not least be relevant for rhetorical practice. In addition, we see a lot of potential in using the internal mechanics and systematics with which Obama realizes his MM and PM pauses (and possibly also other pause types related to audience interaction) to substantiate and further develop (e.g., in a multimodal direction) new phonetic concepts such as the prosodic-constructions idea of Ward & Avila [73].

5. Conclusion

We presented the results of a pilot study that marks the first step towards a better understanding of how pause behavior in public speech triggers speaker charisma and related traits in audiences. Unlike previous studies, our line of research is linguistically and contextually more focused, multimodally oriented, and analyzes not only the time interval of the pause itself but also the phonetic-prosodic embedding on both sides of the pause.

Two questions were addressed here: (1) How does Obama embed pauses phonetically and realize them visually? (2) Do the phonetic/visual means vary with pause location/type of pause?

As for question (1), we observed that Obama's speech features (even compared to pauses in news reading and poetry recitation) exceptionally long pauses whose silent intervals contrast with extremely abrupt pitch and intensity changes right before and after the pause, in particular in post-message (PM) environments. This abruptness is for PM pauses again in contrast with only small pitch and intensity breaks across the silent intervals, unlike for mid-message (MM) pauses. We further found that Obama uses the more facial expressions inside pauses the longer he realizes his pauses, and while such facial communication gestures are rare in MM pauses, they are abundant in PM pauses.

To investigate, in connection with (2), whether and how Obama employs specific phonetic and visual strategies for each pause type/environment, first a perceptual analysis of pause type was performed based on semantic, syntactic and phonetic information, followed by an acoustic analysis to see whether the perceived pause categories have their own unique acoustic-phonetic reality. We found that Obama distinguishes between his acoustic and visual MM and PM pause signals so clearly and systematically that a linear discriminant analysis was able to classify the MM and PM pause types with an accuracy close to 90 % only based on these signals. Even when only pause-internal or pre- or post-pausal features were taken into account, the discrimination accuracy still remained at about 80 %. The particular, precise, and systematic timing properties in and around Obama's pauses are consistent with previous studies in which machine-learning algorithms were able to predict audience reactions merely from Obama's pause durations.

Our data gave us ideas for follow-up perception experiments, which are currently being carried out and in which we systematically vary both the duration of the pause and the acoustic dynamics, especially before the pause. Taken together, the results could provide new impetus for a more effective and quantitatively based pause training in public-speaking coaching. In addition, follow-up experiments will investigate the role of facial communication gestures for perceived pause length and speaker charisma, and we will investigate to what extent long silent pauses can shape the perception of a speaker in a similar way as a lower-pitched voice with regard to the charisma and related traits.

6. References

1. Duez, D. (1997). Acoustic markers of political power. *Journal of Psycholinguistic Research*, 26, 641-654.
2. Huckle, B. (2017). Obama & The Power Of Pausing In A Presentation. Second Nature - The Business Presentation Skills Experts. *URL https://www.secondnature.com.au/blog/obama-the-power-of-pausing/* (last accessed 22 March, 2023)
3. Niebuhr, O. & K. Fischer (2019). Do not hesitate! – Unless you do it shortly or nasally: How the phonetics of filled pauses determine their subjective frequency and perceived speaker performance. *10th Proc. International Conference of Spoken Language Processing (Interspeech), Graz, Austria*, 544-548.
4. Duez, D. (1982). Silent and non-silent pauses in three speech styles. *Language and Speech*, 25(1), 11-28.
5. Šturm. P. & Volín. J. (2023). Occurrence and Duration of Pauses in Relation to Speech Tempo and Structural Organization in Two Speech Genres. *Languages 2023*, 8. 23.
6. Campione. E. & Véronis. J. (2002) A large-scale multilingual study of silent pause duration. *Proc. 1st International Conference of Speech Prosody,. Aix-en-Provence, France,* 199-202.
7. Kowal, S. (1991). *Über die zeitliche Organisation des Sprechens in der Öffentlichkeit: Pausen, Sprechtempo und Verzögerungen in Interviews und Reden von Politikern*. Stuttgart: Huber.
8. Rehn. A. (2016). Every Great Speaker Is a Fantastic Pauser — On Using Pauses and Silences in Public Speaking. URL: *https://medium.com/the-art-of-keynoting* (last accessed 17 Oct, 2023).
9. Esenwein, J.B. & Carnegie, D. (2021). *The art of public speaking*. Berlin: Neobooks.
10. Soorjoo, M. (2012). *Here's the Pitch - How to Pitch Your Business to Anyone, Get Funded, and Win Clients.* Hoboken: Wiley.
11. Biadsy, F., Rosenberg, A., Carlson, R., Hirschberg, J., & Strangert, E. (2008). A cross-cultural comparison of American, Palestinian, and Swedish perception of charismatic speech. *Proc. 4th Speech Prosody Conference Campinas, Brazil,* 579-583.

12. D'Errico, F., Signorello, R., Demolin, D., & Poggi, I. (2013). The perception of charisma from voice: A cross-cultural study. *Proc. IEEE Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, Switzerland,* 552-557.
13. Silber-Varod, V., Niebuhr, O., Schettino, L., & Barbosa, P.A. (2023). Cultural differences of gaps and overlapping speech in political interviews. *Proc. Disfluency in Spontaneous Speech (DiSS) Workshop 2023, Bielefeld, Germany*, 3-7.
14. Agliati, A., Vescovo, A., & Anolli, L. (2005). Conversation patterns in Icelandic and Italian people: Similarities and differences in rhythm and accommodation. In: L. Anolli, S. Duncan Jr., M. S. Magnusson, and G. Riva (eds), *The Hidden Structure of Interaction: From neurons to culture patterns* (pp. 223–235). Amsterdam: IOS Press.
15. Cao, J. (2023). Prosody influence on (im)politeness perception in Chinese-German intercultural communication. *Journal of Politeness Research*, 2023.
16. Michalsky, J. & O. Niebuhr (2019). Myth busted? Challenging what we think we know about charismatic speech. *Acta Universitatis Carolinae Philologica*, 2019(2), 27-56.
17. Brilman, M. & Scherer, S. (2015). A multimodal predictive model of successful debaters or how i learned to sway votes. *Proc. 23rd ACM International Conference on Multimedia, Brisbane, Australia,* 149-158.
18. Antonakis, J., Bastardoz, N., Jacquart, P., & Shamir, B. (2016). Charisma: An ill-defined and ill-measured gift. *Annual Review of Organizational Psychology and Organizational Behavior*, 3, 293-319.
19. Rosenberg, A. & Hirschberg, J. (2009). Charisma perception from text and speech. *Speech Communication,* 51(7), 640-655.
20. Richter, F. (2022). Podcasts - The new mass medium? Statista! URL: *https://www.statista.com/chart/14306/ podcast-adoption/* (last accessed 24 March, 2023).
21. Liu, E. (2020). Most Fortune 500 CEOs are on social media in 2020. Influential Executive. URL: *https://influential executive.com/how-many-fortune-500-ceos-social-media-2020/* (last accessed 24 March, 2023).
22. Dillmann, T. (2022). Bundesverband für Medientraining legt „CEO Digital Video Index" vor. URL: *https://www.pr-journal.de/nachrichten/social-media-web-2-0/28386-bundesver band-fuer-medientraining-legt-ceo-digital-video-index-vor.html* (last accessed 24 March, 2023).
23. Bosker, H.R. (2017). The role of temporal amplitude modulations in the political arena: Hillary Clinton vs. Donald Trump. *Proc. 18th Interspeech Conference, Stockholm, Sweden*, 2228-2232
24. Bosker, H. R. (2021). The contribution of amplitude modulations in speech to perceived charisma. In Weiss, B., Trouvain, J., BarkatWeiss, B., Trouvain, J., Barkat-Defradas, M., & Ohala, J. J. (Eds.). (2020). *Voice Attractiveness: Studies on Sexy, Likable, and Charismatic Speakers* (pp. 165-181). Springer Nature.
25. Strangert, E., & Gustafson, J. (2008). What makes a good speaker? Subject ratings, acoustic measurements, and perceptual evaluations. *Proc. 9th Interspeech Conference, Brisbane, Australia,* 1688-1691.
26. Niebuhr, O., A. Brem, & J. Voße (2016). What makes a charismatic speaker? A computer-based acoustic-prosodic analysis of Steve Jobs' tone of voice. *Computers in Human Behavior*, 64, 366-382.
27. Hamilton, I.A. (2019). Elon Musk just admitted he suffers from self-doubt over his public speaking, a skill critics have hounded him over. URL: *https://africa.businessinsider.com/tech/elon-musk-just-admitted-he-suffers-from-self-doubt-over-his-public-speaking-a-skill/mrrsk9y* (last accessed 17 Oct, 2023).
28. Guyer, J. J., Fabrigar, L. R., Vaughan-Johnston, T. I., & Tang, C. (2018). The counterintuitive influence of vocal affect on the efficacy of affectively-based persuasive messages. *Journal of Experimental Social Psychology*, 74, 161-173.
29. Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of personality and social psychology*, 37(5), 715.
30. Klofstad, C. A. (2016). Candidate voice pitch influences election outcomes. *Political psychology*, 37(5), 725-738.
31. Klofstad, C. A., Anderson, R. C., & Peters, S. (2012). Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women. *Proc. Royal Society B: Biological Sciences*, 279(1738), 2698-2704.
32. Niehof, J. & Niebuhr, O. (2021). What's pitch got to do with it? – The connection between speech prosody and investor-pitch success. *Proc. 1st International Conference on Tone and Intonation (TAI), Sonderborg, Denmark,* 288-292.
33. Jiang, X, & Pell, M. D. (2017). The sound of confidence and doubt. *Speech Communication*, 88, 106-126.
34. Miller, N., Maruyama, G., Beaber, R. J., & Valone, K. (1976). Speed of speech and persuasion. *Journal of personality and social psychology*, 34(4), 615.
35. Woodall, G. W., & Burgoon, J. K. (1983). Talking fast and changing attitudes: A critique and clarification. *Journal of nonverbal behavior*, 8, 126-142.
36. Baker R.E., Baese-Berk M., Bonnasse-Gahot L., et al. (2011). Word durations in non-native English. *Journal of Phonetics*, 39, 1–17.
37. Berger, S., Niebuhr, O., & Peters, B. (2017). Winning over an audience–A perception-based analysis of prosodic features of charismatic speech. *Proc. 43rd Annual Conference of the German Acoustical Society, Kiel, Germany*, 1454-1457.
38. Niebuhr, O. & Skarnitzl, R. (2019). Measuring a speaker's acoustic correlates of pitch – but which? A contrastive analysis based on perceived speaker charisma. *Proc. 19th International Congress of Phonetic Sciences, Melbourne, Australia*, 1-5.
39. Niebuhr, O. (2022). The Kiel Intonation Model - Forms and Functions of German Prosody. In J. Barnes, S. Shattuck-Hufnagel (eds.), *Prosodic Theory and Practice* (pp. 287-318). Cambridge: MIT Press.
40. Niebuhr, O. (2010). On the phonetics of intensifying emphasis in German. *Phonetica*, 67, 170-198.
41. Erickson, D. & Niebuhr, O. (2023). Articulation of prosody and rhythm: Some possible applications to language teaching. *Proc. 13th Nordic Prosody Conference, Sonderborg, Denmark*, 1-45.
42. Banzina, E. (2016). Consonant lengthening for persuasiveness in L1 and L2 English. *International Journal of Applied Linguistics*, 26(3), 403-419.

43. Banzina, E. (2021). Exploring phonetic cues to persuasive oral presentation: A study with British English speakers and English L2 learners. *Language Teaching Research*.
44. Capone, A. (2010). Barack Obama's South Carolina speech. *Journal of Pragmatics*, 42(11), 2964-2977.
45. Navarretta, C. (2017). Prediction of audience response from spoken sequences, speech pauses and co-speech gestures in humorous discourse by Barack Obama. *Proc. 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, Hungary,* 327-332.
46. Strangert, E. (1993). Speaking style and pausing. *PHONUM,* 2, 121–137.
47. Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. *The Journal of the Acoustical Society of America*, 101(1), 514-521.
48. Kohtz, L. S. & Niebuhr, O. (2017). How Long is Too Long? How Pause Features After Requests Affect the Perceived Willingness of Affirmative Answers. *Proc. 18th Interspeech Conference, Stockholm, Sweden*, 3792-3796.
49. Matzinger, T., Pleyer, M., & Żywiczyński, P. (2023). Pause Length and Differences in Cognitive State Attribution in Native and Non-Native Speakers. *Languages 2023*, 8(1), 26.
50. De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior research methods*, 41(2), 385-390.
51. Duez, D. (1993). Acoustic correlates of subjective pauses. *Journal of Psycholinguistic Research*, 22, 21-39.
52. De Jong, N. H. & Bosker, H. R. (2013). Choosing a threshold for silent pauses to measure second language fluency. *Proc. 6th Workshop on Disfluency in Spontaneous Speech (DISS), Bielefeld, Germany*, 17-20.
53. Boersma, P. (2014). The use of Praat in corpus research. In: J. Durand, U. Gut, & G. Kristoffersen (eds), *The Oxford handbook of corpus phonology* (pp. 342-360). Oxford: Oxford University Press.
54. Byrd, D., Krivokapić, J., & Lee, S. (2006). How far, how long: On the temporal scope of prosodic boundary effects. *The Journal of the Acoustical Society of America*, 120(3), 1589-1599.
55. Machač, P. & Skarnitzl, R. (2009). *Principles of phonetic segmentation*. Prague: Epocha.
56. Valstar, M., & Pantic, M. (2006). Fully automatic facial action unit detection and temporal analysis. *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshop, New York, USA*, 149-153.
57. Troille, E., Cathiard, M. A., & Abry, C. (2010). Speech face perception is locked to anticipation in speech production. *Speech Communication*, 52(6), 513-524.
58. Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4), 555-568.
59. Paschen, L., Fuchs., S., & Seifart. F. (2022). Final lengthening and vowel length in 25 languages. *Journal of Phonetics*, 94, 101179.
60. Chen, H. (2006). *Utterance final lengthening and focus induced lengthening in standard Chinese bi-syllabic words.* MA thesis, The University of Edinburgh, Edinburgh, UK.
61. Lehiste, I. (1976). Influence of fundamental frequency pattern on the perception of duration. *Journal of Phonetics*, 4(2), 113-117.
62. Weirich, M. & Simpson, A. P. (2014). Differences in acoustic vowel space and the perception of speech tempo. *Journal of Phonetics*, 43, 1-10.
63. Ohala, J. J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. *Sound symbolism*, 2, 325-347.
64. Turk. A. E. & Shattuck-Hufnagel. S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, 35(4), 445-472.
65. Jäkel, F., Singh, M., Wichmann, F. A., & Herzog, M. H. (2016). An overview of quantitative approaches in Gestalt perception. *Vision research*, 126, 3-8.
66. Pépiot, E. (2014). Male and female speech: a study of mean f0, f0 range, phonation type and speech rate in Parisian French and American English speakers. *Proc. 7th International Conference of Speech Prosody, Dublin, Ireland,* 305-309.
67. Syrdal, A. K. (1996). Acoustic variability in spontaneous conversational speech of American English talkers. *Proc. IEEE 4th International Conference on Spoken Language Processing, Philadelphia, USA*, 438-441.
68. Lang, D. (1992). *The New Secrets of Charisma - How to Discover and Unleash Your Hidden Powers*. New York: SPI Books.
69. Kashino, M. (2006). Phonemic restoration: The brain creates missing speech sounds. *Acoustical Science and Technology*, 27(6), 318-321.
70. Beattie, G. & Aboudan, R. (1994). Gestures, pauses and speech: An experimental investigation of the effects of changing social context on their precise temporal relationships. *Semiotica*, 99(3-4), 239-272.
71. Esposito, A., & Marinaro, M. (2007). What pauses can tell us about speech and gesture partnership. *NATO Security Through Science Series (Human and Societal Dynamics),* 18, 45.
72. McDuff, D., Mahmoud, A., Mavadati, M., Amr, M., Turcot, J., & el Kaliouby, R. (2016). AFFDEX SDK: A Cross-Platform Real-Time Multi-Face Expression Recognition Toolkit. *Proc. 2016 CHI Conference on Human Factors in Computing Systems, New York, USA,* 3723–3726.
73. Ward, N. G. & Avila, J. E. (2023). A dimensional model of interaction style variation in spoken dialog. *Speech Communication*, 149, 47-62.