Chapter 10
The Contribution of Amplitude Modulations in Speech to Perceived Charisma

Hans Rutger Bosker

Abstract Speech contains pronounced amplitude modulations in the 1–9 Hz range, correlating with the syllabic rate of speech. Recent models of speech perception propose that this rhythmic nature of speech is central to speech recognition and has beneficial effects on language processing. Here, we investigated the contribution of amplitude modulations to the subjective impression listeners have of public speakers. The speech from US presidential candidates Hillary Clinton and Donald Trump in the three TV debates of 2016 was acoustically analyzed by means of modulation spectra. These indicated that Clinton's speech had more pronounced amplitude modulations than Trump's speech, particularly in the 1–9 Hz range. A subsequent perception experiment, with listeners rating the perceived charisma of (low-pass filtered versions of) Clinton's and Trump's speech, showed that more pronounced amplitude modulations (i.e., more 'rhythmic' speech) increased perceived charisma ratings. These outcomes highlight the important contribution of speech rhythm to charisma perception.

Keywords Amplitude modulations · Speech rhythm · Modulation spectrum · Charisma perception · Temporal envelope · Political debates

10.1 Introduction

Any spoken utterance, regardless of talker, language, or linguistic content, contains fast-changing spectral information (e.g., vowel formants, consonantal frication, etc.) as well as slower changing temporal information. The temporal information in speech is particularly apparent in the temporal envelope of speech, which includes the fluctuations in amplitude from consonants (constricted vocal tract, lower amplitude) to vowels (unconstricted vocal tract, higher amplitude), from stressed (prominent) to

H. R. Bosker (✉)
Max Planck Institute for Psycholinguistics, P. O. Box 310, 6500 AH Nijmegen, The Netherlands
e-mail: HansRutger.Bosker@mpi.nl

Psychology of Language Department, Donders Institute for Brain Cognition and Behaviour, Radboud University, Kapittelweg 29, 6525 EN Nijmegen, The Netherlands

B. Weiss et al. (eds.), *Voice Attractiveness*, Prosody, Phonology and Phonetics,
https://doi.org/10.1007/978-981-15-6627-1_10

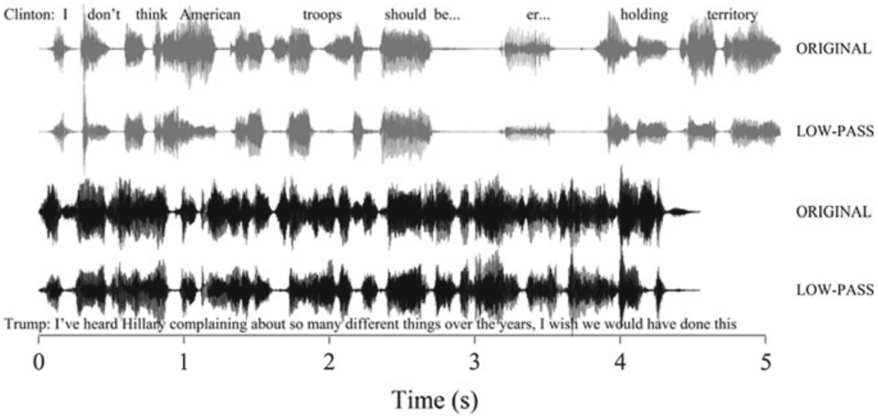

Fig. 10.1 Excerpts of Clinton's speech (in gray) with a notable syllabic rhythm around 3 Hz and Trump's speech (in black) with a notable lack of consistent slow-amplitude modulations. Below each waveform are the low-pass filtered versions of the excerpts, demonstrating that the original slow-amplitude modulations are maintained to a large degree

unstressed syllables (less prominent), etc. For instance, the top example in Fig. 10.1 has pronounced fluctuations in amplitude (also known as amplitude modulations) occurring at around 3 Hz, related to the syllabic rate of the utterance (i.e., roughly three syllables per second).

The temporal dynamics of speech (e.g., energy patterns and syllable durations in speech) are semi-regular at multiple (segmental, syllabic, sentential) timescales (Poeppel, 2003; Rosen, 1992). Hence, speech is an intrinsically rhythmic signal, with 'rhythmic' referring to the semi-regular recurrence over time of waxing and waning prominence profiles in the amplitude signature of speech (for other conceptualizations of speech rhythm, see Kohler, 2009; Nolan & Jeon, 2014). Naturally produced syllable rates typically do not exceed a rate of 9 Hz (Ghitza, 2014; Jacewicz, Fox, & Wei, 2010; Pellegrino, Coupé, & Marsico, 2011; Quené, 2008; Varnet, Ortiz-Barajas, Erra, Gervain, & Lorenzi, 2004). As such, most of the energy in the amplitude modulations in the speech signal is found below 9 Hz (Ghitza & Greenberg, 2009; Greenberg & Arai, 1999, 2004), across a range of typologically distant languages (Ding et al., 2017; Varnet, Ortiz-Barajas, Erra, Gervain, & Lorenzi, 2017), with the most prominent modulation frequencies near the average syllable rate of 3–4 Hz (Delgutte 1998).

In recent models of speech perception (Ghitza 2011; Giraud & Poeppel, 2012; Peelle & Davis, 2012), this rhythmic nature of speech is said to play a central role in speech recognition. For instance, speakers who are intrinsically more intelligible than others show more pronounced low-frequency modulations in the amplitude envelope (Bradlow, Torretta, & Pisoni, 1996). In fact, when the slow amplitude fluctuations in speech are degraded or filtered out, intelligibility drops dramatically (Drullman, Festen, & Plomp, 1994; Ghitza, 2012; Houtgast & Steeneken, 1973), while speech with only minimal spectral information remains intelligible as long as low-frequency temporal modulations are preserved (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). Similarly, speech stream segregation (understanding speech in noise; Aikawa & Ishizuka, 2002), word segmentation (resolving continuous speech into words; Cutler, 1994; Cutler & Butterfield, 1992; Cutler & Norris, 1988), and phoneme perception (Bosker, 2017a; Bosker & Ghitza, 2018; Quené, 2005) are all influenced by regular energy fluctuations in speech.

A powerful demonstration of the contribution of regular amplitude modulations to speech comprehension is the finding that otherwise unintelligible speech can be made intelligible by imposing an artificial rhythm (Bosker & Ghitza, 2018; Doelling, Arnal, Ghitza, & Poeppel, 2014; Ghitza, 2012, 2014). For instance, Bosker and Ghitza (2018) took Dutch recordings of seven-digit telephone numbers (e.g., “215–4653”) and compressed these by a factor of 5 (i.e., make the speech five times as fast while preserving spectral properties such as pitch and formants). This heavy compression manipulation made the intelligibility of the telephone numbers drop from the original 99% to about 39% digits correct. However, Bosker and Ghitza then imposed an artificial rhythm onto the heavily compressed speech, by taking 66 ms windows of compressed speech and spacing these apart by 100 ms of silence (i.e., inserting 100-ms silent intervals). This ‘repackaged’ condition did not contain any additional linguistic or phonetic information compared to the heavily compressed speech; it only differed in having a very pronounced amplitude modulation around 6 Hz. The authors found that imposing this artificial rhythm onto the compressed speech boosted intelligibility (from 39 to 71%) digits correct, demonstrating that regular amplitude modulations play a central role in speech perception.

Rhythmic amplitude modulations in speech not only affect speech intelligibility but they also play a role in spoken communication more generally. For instance, syntactic processing (Roncaglia-Denissen, Schmidt-Kassow, & Kotz, 2013), semantic processing (Rothermich, Schmidt-Kassow, & Kotz, 2012), and recognition memory (Essens & Povel 1985) are all facilitated by regular meter. Moreover, there are even suggestions in the literature that listeners explicitly prefer listening to speech with a clear rhythmic structure. For instance, Obermeier et al. (2013) took four-verse stanzas from old German poetry and independently manipulated the rhyme and meter of these poetry fragments. Rhyme was manipulated by substituting rhyming sentence-final words with non-rhyming words with the same metrical structure (maintaining meter), while meter was manipulated by substituting a sentence-medial word with a word with mismatching metrical structure (e.g., “Nacht” > “Dunkelheit”; maintaining rhyme in sentence-final words). Native German participants rated the original and manipulated fragments of poetry on liking and perceived intensity. Results indicated that non-rhyming and non-metrical stanzas received lower ratings on both the liking and perceived intensity scales, suggesting that the presence of rhythmical structure induces greater esthetic liking and more intense emotional processing (Obermeier et al., 2013, 2016).

Here, we examined the contribution of rhythmic amplitude modulations to the perception of charisma in public speakers' voices. Charisma and charismatic leadership are intensively studied topics, with clear implications for public speakers, politics, religion, and society at large. There seems to be a consensus in the literature that being a charismatic speaker is a necessary precondition for being a charismatic leader. In fact, how one speaks (i.e., performance characteristics, such as pitch, loudness, prosody, etc.) has been argued to contribute to charisma perception more than what one says (i.e., the linguistically formulated communicative message; Awamleh & Gardner, 1999; Rosenberg & Hirschberg, 2009). Several studies have, therefore, attempted to find acoustic correlates of charisma in public speakers' voices (see also in this volume; Rosenberg & Hirschberg, this volume; Brem & Niebuhr, this volume). For instance, pausing behavior (D'Errico, Signorello, Demolin & Poggi, 2013), speech rate (D'Errico et al., & Poggi, 2013), overall intensity (Niebuhr, Voße & Brem, 2016), number and type of disfluencies (Novák-Tót, Niebuhr, & Chen, 2017), and timbre (Weiss and Burkhardt, 2010) have all been identified as contributing to perceived charisma and personality. However, although there are suggestions in the literature that greater variability in pitch and intensity contours increases perceived charisma (D'Errico et al., 2013; Niebuhr et al., 2016; Rosenberg & Hirschberg, 2009), it is unclear what the role of the rhythm of speech is in charisma perception. Therefore, the present research goal was to investigate how political debaters make use of variation in the amplitude envelope in speech production and how this variation, in turn, may affect speech perception.

Regarding rhythm in speech production, we report an acoustic comparison of the temporal amplitude modulations in the speech produced by two presidential candidates in the American elections of 2016: Hillary Clinton and Donald Trump. Recordings from three national presidential debates were collected and the speech produced by both candidates was first matched for overall intensity. Thereafter, their speech was analyzed by means of modulation spectra (Bosker & Cooke, 2018; Ding et al., 2017; Krause & Braida, 2004). These modulation spectra quantify the power of individual modulation frequency components present in a given signal (e.g., see Fig. 10.2), with power on the y-axis and modulation frequency on the x-axis. They can be used to assess which modulation frequencies are most prominent in different signals (e.g., speech and music show well-separated peaks around 5 and 2 Hz, respectively; Ding et al. 2017) but also to compare the overall power (in different frequency bands) across talkers or speech registers (Krause & Braida, 2004). For instance, Bosker and Ghitza 2018 calculated modulation spectra of spoken sentences produced in quiet (plain speech) and the same sentences produced in noise (Lombard speech). Results showed greater power in Lombard speech compared to plain speech, particularly in the 1–4 Hz range, demonstrating that talkers produce more pronounced amplitude modulations when talking in noise, presumably to aid speech comprehension.

Similarly, the present acoustic analysis compared the power of different modulation frequency bands across the two talkers. Greater power in the modulation spectrum of one speaker over another would reveal a more pronounced temporal envelope in that particular candidate's speech (i.e., greater amplitude modulations). Specifically, we expect power differences to occur within the frequency range of typical speech rates, namely below 9 Hz because (1) this modulation range is most characteristic of spontaneous speech (Ding et al., 2017); and (2) previous research indicates that differences between speech registers (plain vs. Lombard speech) are apparent in the lower modulation range (Bosker and Ghitza 2018). Power differences in this 1–9 Hz modulation range would be indicative of a more regular syllabic rhythm. Moreover, the locations of peaks in the modulation spectrum would reveal which modulation frequencies are most pronounced in that speaker's amplitude envelope, being indicative of a specific rhythm preference. By contrast, differences in the power of modulation frequencies between 9–15 Hz are expected to be smaller (if present at all) since this modulation range is less pronounced in speech and is not straightforwardly related to particular acoustic or perceptual units in speech.

When it comes to quantifying rhythm in speech, modulation spectra have several advantages over other rhythm metrics that have been introduced in the literature, such as %V (percentage over which speech is vocalic; Ramus et al. (1999)), $Theta$C (standard deviation of consonantal intervals; Ramus et al. (1999)), PVI (pairwise variability index; Grabe and Low (2002)), or normalized metrics such as VarcoV and VarcoC (Dellwo, 2006; White and Mattys, 2007). These metrics assess durational variability (Loukina et al., 2011), not necessarily periodicity. That is, both isochronous and anisochronous distributions of vowels and consonants can have the same %V. Moreover, such measures are influenced by between-language differences, whereas modulation spectra are not (Ding et al., 2017).

Going beyond merely identifying differences in the use of rhythm between speakers in speech production, we also tested the contribution of pronounced amplitude modulations to speech perception. Specifically, a rating experiment was carried out with low-pass filtered versions of (a subset of) the speech from both speakers. Filtering was applied to reduce the contribution of lexical-semantic information to participants' judgments while maintaining the temporal structure of the acoustic signal (see Fig. 10.1), forcing listeners to base their judgments primarily on temporal characteristics. In line with the introduced beneficial effects of rhythmic regularity on speech intelligibility and esthetic liking, we hypothesized that the perceived charisma ratings would correlate with the speech rhythm in the signals. That is, speech fragments with more pronounced amplitude modulations in the 1–9 Hz range would be expected to be rated as more charismatic than speech fragments with less pronounced amplitude modulations. If corroborated, this would indicate that speech rhythm not only contributes to intelligibility and the qualitative appreciation of the linguistic message but also to the subjective impression listeners have of a (public) speaker.

10.2 Acoustic Analysis

10.2.1 Method

10.2.1.1 Materials

Recordings of all three presidential debates between Hillary Clinton and Donald Trump were retrieved from Youtube. The first debate (NBC News 2016) took place at Hofstra University, Hempstead, NY, USA, on September 26, 2016, and had the form of a traditional debate: the two candidates responded to questions posed by a moderator. The second debate (ABC News, 2016a) was broadcasted from Washington University in St. Louis, St. Louis, MO, USA, on October 9, 2016. This debate was structured as a 'town hall discussion' with the candidates responding mostly to audience member questions. To illustrate, Fig. 10.1 shows two excerpts of Clinton's and Trump's speech in the second debate. The presence of a 3 Hz syllabic 'beat' is clearly visible in Clinton's waveform, whereas Trump's speech notably lacks slow-amplitude modulations. Finally, the third debate (ABC News, 2016b) took place at the University of Nevada, Las Vegas, Las Vegas, NV, USA, on October 19, 2016, and had the form of a traditional debate again.

All monologue speech from either candidate was manually annotated. That is, only those speech fragments in which one talker and one talker alone was speaking (uninterrupted monologue including all pauses, corrections, hesitations, etc.) was analyzed. Speech fragments that included crosstalk, laughter, applause, questions posed by the moderator, etc., were excluded from analyses. Monologues longer than approximately 35 s were cut into smaller fragments of <35 s at sentence boundaries. For the first debate, these annotations resulted in 93 speech fragments produced by Clinton (duration: $M = 24s$; $SD = 7$ s; $range = 5$–36 s; $total = 2263$ s) and 98 speech fragments produced by Trump (duration: $M = 25$ s; $SD = 7$ s; $range = 6$–35 s; $total = 2514$ s). For the second debate, these annotations resulted in 77 speech fragments produced by Clinton (duration: $M = 29$ s; $SD = 5$ s; $range = 8$–36 s; $total = 2243$ s) and 82 speech fragments produced by Trump (duration: $M = 27$ s; $SD = 6$ s; $range = 7$–35 s; $total = 2241$ s). For the third debate, these annotations resulted in 93 speech fragments produced by Clinton (duration: $M = 24$ s; $SD = 7$ s; $range = 5$–35 s; $total = 2245$ s) and 76 speech fragments produced by Trump (duration: $M = 23$ s; $SD = 8$ s; $range = 5$–34 s; $total = 1779$ s).

10.2.1.2 Procedure

Before analysis of the speech fragments, the overall power (root mean square; RMS) in each fragment was normalized (set to an arbitrary fixed value), thus matching the overall power of the speech from both speakers. Following this normalization procedure, the speech fragments from each debate were analyzed separately.

First, the modulation spectrum of each individual speech fragment produced by Clinton was calculated, using a method adapted from (Bosker and Cooke 2018). It involved filtering the speech fragment by a band-pass filter spanning the 500–4000 Hz range and deriving the envelope of the filter's bandlimited output (i.e., Hilbert envelope). The envelope signal was zero-padded to the next power of 2 higher than the length of the longest fragment of that particular speaker to achieve the same frequency resolution across recordings. This signal was then submitted to a Fast Fourier Transform (FFT), resulting in the modulation spectrum of that particular speech fragment. Finally, the average power in two frequency bands was calculated: average power in the 1–9 Hz range and average power in the 9–15 Hz range, resulting in two different observations for each of the speech fragments. Note that natural speech rates typically fall below 9 Hz. The same steps were then repeated for Trump's speech fragments.

This analysis procedure was followed for each of the three debates and formed the two dependent variables (average power below and above 9 Hz) for statistical analyses reported below. In order to visualize the average rhythmicity in the speech of one speaker in one debate, all individual modulation spectra of one speaker in one debate were downsampled by a factor of 25 and thereafter averaged.

10.2.2 Results

Data from the three debates are reported separately to allow for comparison across debates. Note, however, that follow-up analyses did not reveal large qualitative differences between the outcomes of the three debates.

10.2.2.1 First Debate

The average modulation spectra of the speech produced by both speakers in each of the three debates is given in Fig. 10.2.

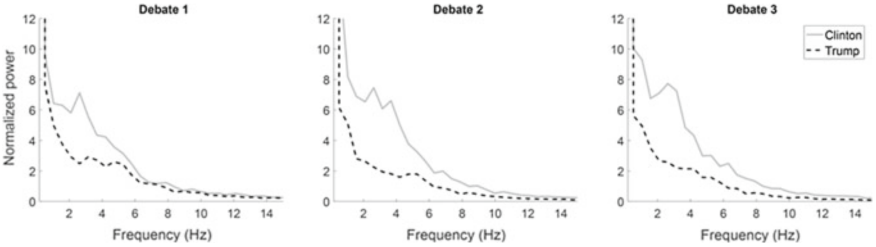

Fig. 10.2 Average modulation spectra of the speech produced by Hillary Clinton (gray solid lines) and Donald Trump (black dashed lines), separately for the three presidential debates

A simple linear model was built in R (R Development Core Team, 2012) separately for each of the two frequency bands (1–9 and 9–15 Hz), predicting the average power for each of the two speakers. The first model, predicting power in the 1–9 Hz range, showed a significant effect of Speaker ($b = 1.265$, $F(1, 189) = 90.91$, $p < 0.001$, $adjustedR^2 = 0.321$), indicating that Clinton's speech contained more power in the lower frequencies compared to Trump's speech. The other model, predicting power in the 9–15 Hz range, also showed a significant difference between the two speakers, only with a much smaller effect size ($b = 0.164$, $F(1, 189) = 42.75$, $p < 0.001$, $adjusted\, R^2 = 0.180$). These findings reveal that, in the first presidential debate, Clinton's speech contained more power in the 1–9 Hz range, and also slightly more power in the frequency range above 9 Hz.

10.2.2.2 Second Debate

The average modulation spectra of all speech produced by the two speakers in the second debate are given in Fig. 10.2.

Again, simple linear models were built separately for each of the two frequency bands (1–9 Hz and 9–15 Hz). The first model, predicting power in the 1–9 Hz range, showed a significant effect of Speaker ($b = 2.322$, $F(1, 157) = 434.5$, $p < 0.001$, $adjustedR^2 = 0.733$), as did the second model, predicting power in the 9–15 Hz range, only with a considerably smaller effect size ($b = 0.263$, $F(1, 157) = 250.9$, $p < 0.001$, $adjusted\, R2 = 0.613$). These findings reveal that, in the second presidential debate, Clinton's speech contained considerably more power in the 1–9 Hz range, and also somewhat more power in the frequency range above 9 Hz.

Note that, similar to the first debate, there is a clear peak in the modulation spectrum of Clinton around 3 Hz. This peak indicates a pronounced syllabic rhythm around 3 Hz in the amplitude envelope of Clinton's speech (cf. Fig. 10.1).

10.2.2.3 Third Debate

The average modulation spectra of the speech produced by both speakers in the third debate are given in Fig. 10.2.

Once more, simple linear models were built separately for each of the two frequency bands (1–9 Hz and 9–15 Hz). The first model, predicting power in the 1–9 Hz range, showed a significant effect of Speaker ($b = 2.427$, $F(1, 167) = 207.5$, $p < 0.001$, $adjusted\, R^2 = 0.551$), as did the second model, predicting power in the 9–15 Hz range, only with a considerably smaller effect size ($b = 0.350$, $F(1, 167) = 197.6$, $p < 0.001$, $adjusted\, R^2 = 0.539$). These findings from the third debate mirror those from the second debate: Clinton's speech contained considerably more power in the 1–9 Hz range, and also slightly more power in the frequency range above 9 Hz.

10.3 Perception Experiment

10.3.1 Participants

Native Dutch participants ($N = 20$; 17 females, 3 males; $M_{age} = 25$) with normal hearing were recruited from the Max Planck Institute's participant pool. Participants in all experiments reported here gave informed consent as approved by the Ethics Committee of the Social Sciences department of Radboud University (project code: ECSW2014-1003-196).

10.3.2 Material

Only speech fragments from the third debate were included in the perception experiment because (1) it was impossible to include the speech from all debates in a single rating experiment for reasons of length and (2) the third debate showed the largest difference between the two talkers in the power of amplitude modulations in the 1–9 Hz range.

Speech fragments from the third debate were first scaled to 70 dB using Praat (Boersma & Boersma, 2016). We did not want raters to base their judgments on the linguistic content of the speech since this was not controlled across the two speakers. Therefore, all speech was low-pass filtered (450 Hz cutoff, using a Hann window with a roll-off width of 25 Hz as implemented in Praat) to avoid lexical-semantic interference, while preserving sufficient ecological validity (being like naturally filtered speech, as if overhearing a person in another room). This manipulation crucially leaves the amplitude fluctuations present in the original speech signals relatively intact (cf. Fig. 10.1). After low-pass filtering, the speech was scaled to 70 dB.

10.3.3 Procedure

Participants in the experiment listened to the low-pass filtered speech fragments from either Clinton or Trump (counter-balanced across participants) in random order. Participants were instructed to rate the items for charisma, basing their judgments on the sound of the speech. They were explicitly pointed to the speaker's identity (but remained unaware that ratings of the other speaker were also collected). Nevertheless, they were told not to let any potential political or personal preferences influence their ratings. The use of a between-participants design reduced the contrast between the two speakers, thus further minimizing potential biases due to speaker sex, pitch, political stance, etc. Participants were instructed to rate the items for charisma using an Equal Appearing Interval Scale (Thurstone, 1928), including seven stars with labeled extremes (not charismatic on the left; very charismatic on the right).

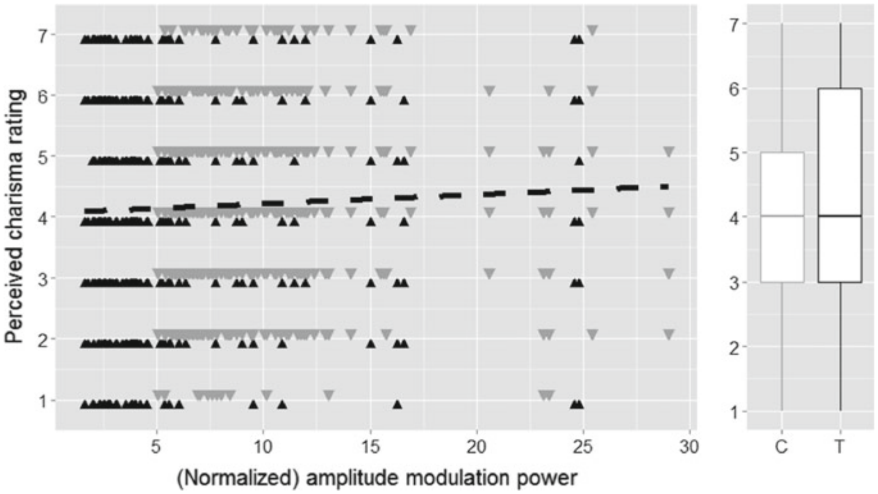

Fig. 10.3 *Left panel*: Individual perceived charisma ratings (on a scale from 1 "not charismatic" to 7 "very charismatic") of each speech fragment as a function of the (normalized) average power of amplitude modulations in the 1–9 Hz range. Gray triangles indicate speech fragments from Clinton and black triangles those from Trump. The black dashed line shows a (simple) linear regression line across all data points. *Right panel:* Boxplots showing the charisma ratings split for the two speakers (C = Clinton; T = Trump)

10.3.4 Results

The average perceived charisma rating of the speech of Clinton was 4.1, while Trump received an average rating of 4.3. Speech fragments with outlier values for the average power of amplitude modulations in the 1–9 Hz range (i.e., $> 2 * SD$; $n = 8$) were excluded to avoid the heavy weight of these outliers on the correlation analyses reported below. Figure 10.3 shows the individual perceived charisma ratings of speech fragments as a function of the average power of amplitude modulations in the 1–9 Hz range.

The right panel of Fig. 10.3 suggests that, on average, Trump (black) received higher charisma ratings than Clinton (gray). The left panel suggests that the charisma ratings seem to be a function of the average power of amplitude modulations in the 1–9 Hz range, with greater power of the amplitude modulations leading to higher charisma ratings.

Perceived charisma ratings were entered into a simple linear model, including the predictor's Speaker (categorical predictor; deviation coding, with Trump coded as -0.5 and Clinton as $+0.5$), Modulation Power Below 9 Hz (continuous predictor; z-scored), Modulation Power Above 9 Hz (continuous predictor; z-scored), and interactions between Speaker and the two Modulation Power predictors. This model, first, revealed a significant effect of Modulation Power Below 9 Hz ($b = 0.318$, $F(5, 1664) = 2.245$, $p = 0.041$). This indicates that, across the two talkers, speech with greater power in the 1–9 Hz range led to higher charisma ratings. Second, we found a main effect of Speaker ($b = -0.209$, $F(5, 1664) = 2.245$, $p = 0.014$), suggesting that Trump's speech was rated as more charismatic overall than Clinton's speech. No effect of Modulation Power Above 9 Hz was observed (p = 0.151), nor was their statistical evidence for either interaction term.

10.3.5 General Discussion

The present research goal was to investigate the role of temporal amplitude modulations in charisma perception in political debates. An acoustic analysis of the speech from two presidential candidates, Hillary Clinton and Donald Trump, in three different debates was carried out by means of modulation spectra, revealing the spectral content of the amplitude envelopes. Also, a perception experiment investigated whether judgments of perceived charisma would be sensitive to the speech rhythm in the acoustic signal.

Comparison of the amplitude spectra of Hillary Clinton's and Donald Trump's speech revealed considerably greater power in the modulation spectra of Clinton's speech than in those of Trump's speech. This power difference cannot be due to overall intensity differences between the two speakers since all speech was normalized in overall power prior to analysis, matching the overall intensity of Clinton's and Trump's speech fragments. Also, the power difference cannot be attributed to differences in habitual speech rate since such differences would be expected to lead to peaks at different frequencies in the modulation spectra, rather than differences in overall power. Instead, this finding indicates that there was a more pronounced temporal envelope in Clinton's speech (compared to Trump's speech).

Note that this power difference was concentrated (i.e., largest) in the 1–9 Hz range, the range of typical syllable rates (Ding et al., 2017; Ghitza & Greenberg, 2009; Greenberg & Arai, 1999, 2004). This suggests that the power difference between Clinton and Trump is driven by more pronounced syllabic amplitude fluctuations in the speech of Clinton. Moreover, across the three debates, there seems to be a relatively consistent peak around 3 Hz in Clinton's modulation spectra, suggesting a preferred syllabic rate. In contrast, Trump's modulation spectra lack pronounced peaks, indicating particularly flat, that is, unmodulated amplitude envelope contours.

Whether or not Clinton used this particular speaking style (with regular amplitude modulations) purposefully and strategically remains unknown. In this regard, one may note that speakers, in general, tend to produce greater amplitude modulations when instructed to produce clear speech (Krause & Braida, 2004) or when talking in noise (Bosker & Cooke, 2018), presumably for reasons of achieving greater intelligibility. As such, Clinton's speaking style during the three debates examined here may be the result of her extensive experience with making herself understood during public addresses. We may speculate that the influence of the enhanced modulation signature of Clinton's speech did not influence charisma perception alone. Regular energy fluctuations have been shown to benefit speech recognition (Doelling et al., 2014; Ghitza, 2012, 2014), particularly in noisy listening conditions (Aikawa and Ishizuka, 2002), and, as such, may have improved Clinton's intelligibility in the noisy environment of a live debate. This seems particularly relevant considering the large number of interruptions (i.e., overlapping speech) that Clinton encountered during the three debates (Trump: $N = 106$ vs. Clinton : $N = 27$). Also, rhythmic amplitude modulations facilitate recognition memory (Essens & Povel 1985), potentially serving Clinton's political aims at the time.

One may also speculate about the absence of amplitude modulations in Trump's speech. Tian's recent analysis (Tian, 2017) of Trump's disfluency patterns during these presidential debates indicated that Trump was considerably more disfluent than Clinton. Trump was found to use particularly many repetitions, repairs, and abandoned utterances (Tian, 2017); all types of disfluencies that signal less extensive utterance planning and self-monitoring. As such, Tian suggested that Trump used less rehearsed utterances compared to Clinton. This difference in utterance planning can well be thought to underlie the difference in rhythmic structure between the two speakers: putting more effort in cognitive planning would also allow the speaker to better temporally organize the syllabic structure of the utterance, and especially so with increased public-speaking experience.

The outcomes of the perception experiment supported two conclusions. First, more pronounced amplitude modulations biased raters toward higher perceived charisma ratings. Across all speech fragments from both talkers, we observed that those items with a higher power of amplitude modulations in the 1–9 Hz range also received higher perceived charisma ratings—independent from the main speaker effect. This suggests that the rhythm of speech contributes to perceived charisma, with implications for public speakers in general.

The second conclusion is that Trump's speech was, on the whole, rated as more charismatic than Clinton's. Although this may seem at odds with the observation that less pronounced amplitude modulations result in lower perceived charisma ratings, it is important to realize that listeners could base their judgments on a larger set of acoustic characteristics than just rhythm. It is unlikely that participants in the study based their perceived charisma ratings solely on the amplitude modulation signatures of the speech signals. Many other (acoustic) characteristics are likely to have contributed to participants' judgments—even in the case of low-pass filtered speech (i.e., without access to linguistic content). One potential acoustic cue that was available to listeners and that may account for the main effect of Speaker is pitch. The low-pass filter applied to the speech only filtered out spectral information above 450 Hz, leaving fundamental frequencies relatively intact. As such, the low-pass filtered stimuli still contained acoustic cues to talker gender (distinction male vs. female cued by pitch). Indeed, talker gender is known to bias charisma ratings (and the perception of other personality traits), with male talkers generally being perceived as more charismatic than female talkers (Brooks, Huang, Kearney, & Murray, 2014; Niebuhr, Skarnitzl, & Tylecková, 2018; Novák-Tát, 2017). Therefore, the main effect of Speaker is likely driven by a range of acoustic and social factors that were not controlled for. Still, it is important to note that the correlation between more pronounced amplitude modulations and higher perceived charisma ratings held across talkers (no interaction between modulation power and speaker). This means that, despite an overall difference between the male and female voice, enhanced amplitude modulations in speech equally affected the ratings of Trump's and Clinton's speech.

Another possible explanation for the overall effect of Speaker could be related to the concept of 'effectiveness windows' in charisma perception (Niebuhr, Tegtmeier, & Brem, 2017). It has been proposed that public speakers, in attempting to persuade their audiences, should use charisma-relevant acoustic cues within particular functional ranges, avoiding, for instance, exaggerated vocal characteristics. Maybe Clinton's consistent use of regular amplitude modulations was perceived as an "overdose" of charismatic vocal cues, thus at some point hurting, rather than serving, the subjective impression listeners had of her. However, such an interpretation would also predict an inverse U-curve in the relationship between modulation power and charisma perception, such that greater rhythmicity would be beneficial only up to a certain point. However, follow-up statistical analyses (i.e., testing for a quadratic effect of Modulation Power Below 9 Hz) and visual inspection of Fig. 10.3 do not support the presence of such a U-shaped relationship, arguing against this particular explanation.

The fact that we used low-pass filtered speech may be seen as both a strength as well as a limitation of the current study. It is a strength of the methodology of the experiment because this allowed us to isolate the (temporal) acoustics of the speech from the linguistic content. In this fashion, potential interference from the linguistic message was reduced. At the same time, one may argue that it limits the generalizability of the present findings since in most natural communicative situations we hear unfiltered speech. For our current purposes, we valued experimental control higher than ecological validity and future studies may investigate whether the rhythm of speech also influences charisma perception in more natural settings.

Another limitation of this study is that we only performed correlational analyses. Even though we are unaware of possible confounds, we acknowledge that the present empirical evidence does not necessarily warrant the conclusion that more pronounced amplitude modulations causally influence perceived charisma. Future investigations may, for instance, examine this causal relationship by directly manipulating the modulation depth of speech fragments—while keeping all other (acoustic, linguistic, social) cues present in the signal constant.

Finally, one further highly relevant issue in the field of charisma research is the role of listener variation in charisma perception. Most empirical studies of charisma perception have used subjective ratings collected from young university students. In fact, some studies, like the present one, recruited non-native speakers of the language under study (e.g., Brem & Niebuhr, this volume). It remains unclear how variation among raters might impact charisma perception and the perceptual weight assigned to various vocal characteristics. Is charisma perception language- or culture-dependent (cf. D'Errico, 2013)? Do non-native speakers of a language weight the acoustic cues to charisma differently from native speakers, possibly through influences from their L1? Do male and female raters differ in how they judge male versus female public speakers (cf. Brem & Niebuhr, this volume)? What is the role of one's own speech production patterns on the perception of others (cf. Bosker, 2017b)? For instance, do fast talkers find fast speech more attractive or persuasive than others? These questions regarding inter-individual variation in charisma perception are promising avenues for future research.

10.4 Conclusion

The present outcomes shed light on the use and function of speech rhythm in political debates, specifically comparing the speech produced by Hillary Clinton and Donald Trump in three presidential debates in 2016. Clinton's speech was observed to contain more power in the modulation spectra, particularly in the 1–9 Hz range, suggesting more pronounced amplitude modulations in her speech (compared to Trump). This may be argued to indicate that Clinton planned her utterances more extensively, allowing more opportunity to temporally organize the syllabic structure of her utterances. At the same time, the lack of rhythmic amplitude modulations in Trump's speech may indicate a level of spontaneity in his speech production, with little attempt to pre-plan certain utterances.

Perceptual data revealed a positive correlation between the strength of amplitude modulations in the syllabic range (1–9 Hz), on the one hand, and perceived charisma ratings, on the other hand. This suggests that greater rhythm in the speech of a public speaker positively influences listeners' impressions of the speaker charisma. Thus, it highlights the important contribution of speech rhythm to charisma perception.

Acknowledgments The author was supported by a Gravitation grant from the Dutch Government to the Language in Interaction Consortium. Parts of the acoustic analysis have been presented at Interspeech 2017, Stockholm, Sweden. Thanks go to YouTube and the various news agencies for making the digital recordings of the presidential debates freely available. Thanks also to Annelies van Wijngaarden for coordinating the perception experiment, to Joe Rodd for help with visualizing the data, and to the student-assistants in the Psychology of Language Department of the Max Planck Institute for Psycholinguistics for help with annotating the speech recordings.

References

ABC News. (2016). *FULL VIDEO: Donald Trump vs Hillary Clinton—2nd Presidential Debate*. Retrieved October 9, 2016, https://www.youtube.com/watch?v=h-gkBUbU_F4.

ABC News (2016). *FULL VIDEO: Donald Trump vs Hillary Clinton—3rd Presidential Debate*. Retrieved October 9, 2016, from https://www.youtube.com/watch?v=LsA6Gj8y8rU.

Aikawa, K., & Ishizuka, K. (2002). Noise-robust speech recognition using a new spectral estimation method "PHASOR". In *Proceedings of Acoustics, Speech, and Signal Processing (ICASSP)* (pp. 397–400).

Awamleh, R., & Gardner, W. L. (1999). Perceptions of leader charisma and effectiveness: The effects of vision content, delivery, and organizational performance. *The Leadership Quarterly*, *10*(3), 345–373.

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer. Computer program.

Bosker, H. R. & Cooke, M. (2018). Talkers produce more pronounced amplitude modulations when speaking in noise. *Journal of the Acoustical Society of America, 143*(2), EL121-EL126.

Bosker, H. R. (2017a). Accounting for rate-dependent category boundary shifts in speech perception. *Perception & Psychophysics*, *79*(1), 333–343.

Bosker, H. R. (2017b). How our own speech rate influences our perception of others. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(8), 1225–1238.

Bosker, H. R., & Ghitza, O. (2018). Entrained theta oscillations guide perception of subsequent speech: Behavioural evidence from rate normalisation. *Language, Cognition and Neuroscience*, *33*(8), 955–967.

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, *20*(3–4), 255–272.

Brooks, A. W., Huang, L., Kearney, S. W., & Murray, F. E. (2014). Investors prefer entrepreneurial ventures pitched by attractive men. *Proceedings of the National Academy of Sciences*, *111*(12), 4427–4431.

Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua*, *92*, 81–104.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*(2), 218–236.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 113–121.

Delgutte, B., Hammond, B., & Cariani, P. (1998). Neural coding of the temporal envelope of speech: relation to modulation transfer functions. Psychophysical and physiological advances in hearing, 595–603.

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for Δ C. *Language and language-processing* (pp. 231–241). Frankfurt a. M.: Peter Lang.

D'Errico, F., Signorello, R., Demolin, D., & Poggi, I. (2013). The perception of charisma from voice: A cross-cultural study. In *Proceedings of Affective Computing and Intelligent Interaction (ACII)* (552–557).

Ding, N., Patel, A., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience and Biobehavioral Reviews*, *14*(1), 113–121.

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, *85*, 761–768.

Drullman, R., Festen, J. M., & Plomp, R. (1994). Effect of reducing slow temporal modulations on speech recognition. *The Journal of the Acoustical Society of America*, *95*(5), 2670–2680.

Essens, P. J., & Povel, D.-J. (1985). Metrical and nonmetrical representations of temporal patterns. Perception & Psychophysics, 37(1), 1–7. https://doi.org/10.3758/bf03207132

Ghitza, O. (2011). Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology, 2*, 130.

Ghitza, O. (2012). On the role of theta-driven syllabic parsing in decoding speech: Intelligibility of speech with a manipulated modulation spectrum. *Frontiers in Psychology, 3*, 238.

Ghitza, O. (2014). Behavioral evidence for the role of cortical Θ oscillations in determining auditory channel capacity for speech. *Frontiers in Psychology, 5*, 652.

Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, *66*(1–2), 113–126.

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Laboratory Phonology*, *7*, 515–546.

Greenberg, S., & Arai, T. (2004). What are the essential cues for understanding spoken language? *IEICE Transactions on Information and Systems, E87-D*(5), 1059–1070.

Greenberg, S., & Arai, T. (1999). Speaking in shorthand—A syllable-centric perspective for understanding pronunciation variation. *Speech Communication*, *29*(2), 159–176.

Houtgast, T., & Steeneken, H. J. (1973). Modulation transfer-function in room acoustics as a predictor of speech intelligibility. *Acustica*, *28*(1), 66–73.

Jacewicz, E., Fox, R. A., & Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *The Journal of the Acoustical Society of America*, *128*(2), 839–850.

Kohler, K. J. (2009). Rhythm in speech and language. *Phonetica*, *66*(1–2), 29–45.

Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America*, *115*(1), 362–378.

Loukina, A., Kochanski, G., Rosner, B., Keane, E., & Shih, C. (2011). Rhythm measures and dimensions of durational variation in speech. *The Journal of the Acoustical Society of America*, *129*(15), 3258–3270.

NBC News. (2016). *FULL VIDEO: The First Presidential Debate: Hillary Clinton and Donald Trump (Full Debate)*. Retrieved September 28, 2016, from https://www.youtube.com/watch?v=855Am6ovK7s.

Niebuhr, O., Skarnitzl, R., & Tylecková, L. (2018). The acoustic fingerprint of a charismatic voice - Initial evidence from correlations between long-term spectral features and listener ratings. In *Proceedings of Speech Prosody* (pp. 359–363).

Niebuhr, O., Voße, J., & Brem, A. (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.

Niebuhr, O., Tegtmeier, S., & Brem, A. (2017). Advancing research and practice in entrepreneurship through speech analysis—from descriptive rhetorical terms to phonetically informed acoustic charisma metrics. *Journal of Speech Sciences*, *6*(3), 3–26.

Nolan, F., & Jeon, H.-S. (2014). Speech rhythm: A metaphor? *Philosophical Transactions of the Royal Society B-Biological Sciences, 369*(1658).

Novák-Tót, E., Niebuhr, O., & Chen, A. (2017). A gender bias in the acoustic-melodic features of charismatic speech? In *Proceedings of Interspeech* (pp. 2248–2252).

Obermeier, C., Menninghaus, W., von Koppenfels, M., Raettig, T., Schmidt-Kassow, M., Otterbein, S., & Kotz, S. A. (2013). Aesthetic and emotional effects of meter and rhyme in poetry. *Frontiers in Psychology, 4*(10).

Obermeier, C., Kotz, S. A., Jessen, S., Raettig, T., von Koppenfels, M., & Menninghaus, W. (2016). Aesthetic appreciation of poetry correlates with ease of processing in event-related potentials. *Cognitive, Affective, & Behavioral Neuroscience*, *16*(2), 362–373.

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology, 3*(10).

Pellegrino, F., Coupé, C., & Marsico, E. (2011). Across-language perspective on speech information rate. *Language*, *87*(3), 539–558.

Peter, J., & Povel, D. -J. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception & Psychophysics*, *37*(1), 1–7.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication*, *41*(1), 245–255.

Quené, H., & Port, R. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica*, *62*(1), 1–13.

Quené, H. (2008). Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *The Journal of the Acoustical Society of America*, *132*(2), 1104–1113.

R Development Core Team. (2012). R: A Language and Environment for Statistical Computing. Computer program.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*, 265–292.

Roncaglia-Denissen, M. P., Schmidt-Kassow, M., & Kotz, S. A. (2013). Speech rhythm facilitates syntactic ambiguity resolution: ERP evidence. *PloS One*, *8*(2), e56000.

Rosen, S. (1992). Temporal information in speech—acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, *336*(1278), 367–373.

Rosenberg, A., & Hirschberg, J. (2009). Charisma perception from text and speech. *Speech Communication*, *51*(7), 640–655.

Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia*, *50*(2), 232–244.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*(5234), 303.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, *33*, 529–554.

Tian, Y. (2017). Disfluencies in Trump and Clinton first presidential debate. In *Proceedings of the conference Fluency and Disfluency Across Languages and Language Varieties* (pp. 106–109).

Varnet, L., Ortiz-Barajas, M. C., Erra, R. G., Gervain, J., & Lorenzi, C. (2017). A cross-linguistic study of speech modulation spectra. *The Journal of the Acoustical Society of America*, *142*(4), 1976–1989.

Verhoeven, J., De Pauw, G., & Kloots, H. (2004). Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. *Language and Speech*, *47*(3), 279–308.

Weiss, B., & Burkhardt, F. (2010). Voice attributes affecting likability perception. In *Proceedings of Interspeech* (pp. 2014–2017).

White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, *35*(4), 501–522.