

The role of temporal amplitude modulations in the political arena: Hillary Clinton vs. Donald Trump

Hans Rutger Bosker^{1,2}

¹Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH, Nijmegen, Netherlands

²Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands

HansRutger.Bosker@mpi.nl

Abstract

Speech is an acoustic signal with inherent amplitude modulations in the 1-9 Hz range. Recent models of speech perception propose that this rhythmic nature of speech is central to speech recognition. Moreover, rhythmic amplitude modulations have been shown to have beneficial effects on language processing and the subjective impression listeners have of the speaker. This study investigated the role of amplitude modulations in the political arena by comparing the speech produced by Hillary Clinton and Donald Trump in the three presidential debates of 2016.

Inspection of the modulation spectra, revealing the spectral content of the two speakers' amplitude envelopes after matching for overall intensity, showed considerably greater power in Clinton's modulation spectra (compared to Trump's) across the three debates, particularly in the 1-9 Hz range. The findings suggest that Clinton's speech had a more pronounced temporal envelope with rhythmic amplitude modulations below 9 Hz, with a preference for modulations around 3 Hz. This may be taken as evidence for a more structured temporal organization of syllables in Clinton's speech, potentially due to more frequent use of preplanned utterances. Outcomes are interpreted in light of the potential beneficial effects of a rhythmic temporal envelope on intelligibility and speaker perception.

Index Terms: amplitude envelope, amplitude modulations, speech rhythm, modulation spectrum, Hillary Clinton, Donald Trump.

1. Introduction

Speech is an intrinsically rhythmic signal as a consequence of the fact that energy patterns and syllable durations in speech are constrained by the articulatory dynamics of the lips, jaw, and tongue. For instance, naturally produced syllable rates typically do not exceed a rate of 9 Hz [1-5] and, as such, most of the energy in the amplitude modulations in the speech signal is found below 9 Hz [6-8], across a range of typologically distant languages [9].

In recent models of speech perception [10-12], this rhythmic nature of speech has been suggested to play a central role in online speech recognition. For instance, speakers who are intrinsically more intelligible than others show more pronounced low-frequency modulations in the amplitude envelope [13]. In fact, when the slow amplitude fluctuations in speech are destroyed or filtered out, intelligibility drops dramatically [14]. Conversely, imposing an artificial rhythm on extremely impoverished speech can restore intelligibility [2, 14, 15]. Likewise, speech stream segregation

(understanding speech in noise; [16]), word segmentation (resolving continuous speech into words; [17-19]), and phonetic perception [20, 21] are all influenced by regular energy fluctuations in speech.

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 [22], semantic processing [23], and recognition memory [24] are all facilitated by regular meter. Moreover, listeners explicitly prefer listening to speech with clear rhythmic structure as it induces greater aesthetic 'liking' and more intense emotional processing [25, 26]. Finally, greater pitch and intensity modulations correlate with greater perceived charisma in public speakers [27-29].

Given these beneficial effects of (semi-)regular temporal amplitude modulations *in speech perception*, this study investigated how political debaters may use variation in the amplitude envelope *in speech production*. In political debates, speakers purposefully try to convince the audience of their own political views. In such communicative situations, not only *what* is said but also *how* it is said contributes to the speaker's overall communicative success. As such, variation in the amplitude envelope of spoken utterances may be used by politicians to improve their intelligibility, comprehensibility, and even overall 'liking' by the audience.

This paper reports a 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 [9], showing the spectral content of the amplitude envelope of the two speakers. 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. The locations of peaks in the modulation spectrum would reveal which modulation frequencies are most pronounced in that speaker's amplitude envelope.

When it comes to quantifying rhythmicity 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; [30]), ΔC (standard deviation of consonantal intervals; [30]), *PVI* (pairwise variability index; [31]), or normalized metrics such as *VarcoV* and *VarcoC* [32, 33]. These metrics assess durational variability [34]; 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 [9].

2. Acoustic analysis

2.1. Method

2.1.1. Materials

Recordings of all three presidential debates between Hillary Clinton and Donald Trump were retrieved from Youtube. The first debate [35] 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 [36] 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. Finally, the third debate [37] took place at University of Nevada, Las Vegas, Las Vegas, NV, USA, on October 19, 2016, and had the form of a traditional debate again.

All speech from either candidate produced without interruptions (from the other candidate, the moderator, or someone else) was manually annotated. Only uninterrupted monologue from the two candidates was analyzed; all other sounds in the recordings were excluded from analyses (e.g., crosstalk, laughter, applause, questions posed by the moderator, etc.). For the first debate, these annotations resulted in 55 speech fragments produced by Clinton (duration: $M = 42$ s; $SD = 44$ s; $range = 1-167$ s; $total = 2290$ s) and 56 speech fragments produced by Trump (duration: $M = 45$ s; $SD = 43$ s; $range = 2-158$ s; $total = 2532$ s). For the second debate, these annotations resulted in 24 speech fragments produced by Clinton (duration: $M = 93$ s; $SD = 45$ s; $range = 8-170$ s; $total = 2243$ s) and 37 speech fragments produced by Trump (duration: $M = 61$ s; $SD = 47$ s; $range = 1-162$ s; $total = 2257$ s). For the third debate, these annotations resulted in 48 speech fragments produced by Clinton (duration: $M = 48$ s; $SD = 38$ s; $range = 2-126$ s; $total = 2296$ s) and 49 speech fragments produced by Trump (duration: $M = 37$ s; $SD = 37$ s; $range = 1-146$ s; $total = 1810$ s).

2.1.2. Analysis procedure

Before analysis of the speech fragments, the overall power (root-mean-square; RMS) in each fragment was normalized, 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 Krause and Braida [38]. 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, the power in all modulation spectra of one speaker in one debate were downsampled by a factor of 50 and thereafter averaged.

2.2. Results

2.2.1. First debate

The average modulation spectra of the speech produced by both speakers in the first debate are given in Figure 1.

A simple linear model was built separately for each of the two frequency bands (1-9 Hz 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 ($b = .091$, $F(1, 109) = 4.211$, $p = .043$, $adjusted R^2 = .028$), 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, did not show a significant difference between the two speakers ($p = .171$). These findings reveal that, in the first presidential debate, Clinton’s speech contained more power in the 1-9 Hz range.

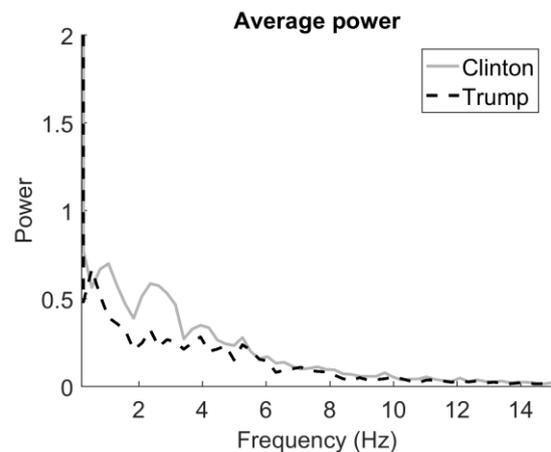


Figure 1: Average modulation spectra of the speech produced by Hillary Clinton (gray solid line) and Donald Trump (black dashed line) in the first presidential debate.

2.2.2. Second debate

Short excerpts of both speakers in the second presidential debate are shown in Figure 2. The average modulation spectra of all speech produced by the two speakers in the second debate are given in Figure 3.

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 ($b = .420$, $F(1, 59) = 60.730$, $p < .001$, $adjusted R^2 = .498$), as did the second model, predicting power in the 9-15 Hz range, only with a considerably smaller effect size ($b = .052$, $F(1, 59) = 40.640$, $p < .001$, $adjusted R^2$

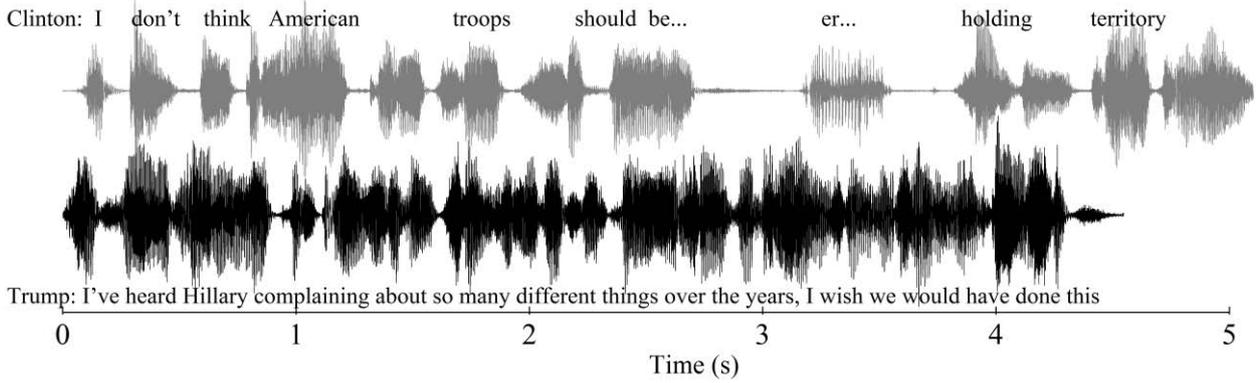


Figure 2: *Top panel: excerpt of Clinton’s speech from the second debate (in gray) with a notable syllabic rhythm around 3 Hz. Bottom panel: excerpt of Trump’s speech from the second debate (in black) with a notable lack of consistent slow amplitude modulations.*

= .398). 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. In Figure 2, showing examples of Clinton’s and Trump’s speech (taken from the second debate), the presence of a 3 Hz ‘beat’ is clearly visible in Clinton’s waveform, whereas Trump’s speech notably lacks slow amplitude modulations.

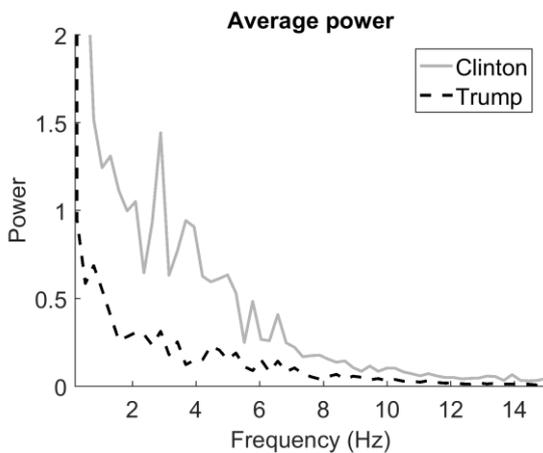


Figure 3: *Average modulation spectra of the speech produced by Hillary Clinton (gray solid line) and Donald Trump (black dashed line) in the second presidential debate.*

2.2.3. Third debate

The average modulation spectra of the speech produced by both speakers in the third debate are given in Figure 4.

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 ($b = .472, F(1, 95) = 41.730, p < .001, adjusted R^2 = .298$), as did the second model, predicting power in the 9-15 Hz range, only with a considerably smaller

effect size ($b = .067, F(1, 95) = 40.910, p < .001, adjusted R^2 = .294$). 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 somewhat more power in the frequency range above 9 Hz.

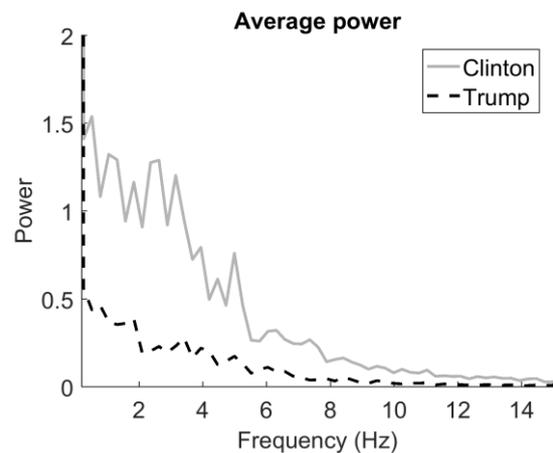


Figure 4: *Average modulation spectra of the speech produced by Hillary Clinton (gray solid line) and Donald Trump (black dashed line) in the third presidential debate.*

3. Discussion

This study investigated the role of temporal amplitude modulations in political debates. Speech from two presidential candidates, Hillary Clinton and Donald Trump, in three different debates was collected and analyzed by means of modulation spectra, revealing the spectral content of the amplitude envelopes.

Comparison of the amplitude spectra of Hillary Clinton’s and Donald Trump’s speech revealed considerably greater power in the modulation spectrum of Clinton’s speech than in that 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 due 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 [6-9]. 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 strong peaks, indicating particularly flat, that is, unmodulated amplitude envelope contours.

Regular energy fluctuations have been shown to benefit speech recognition [2, 14, 15], particularly in noisy listening conditions [16], and, as such, may have improved Clinton's intelligibility in the noisy environment of a live debate. This seems particularly relevant if one considers 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 [24], induce greater 'liking', and correlate with perceived charisma in public speakers [25-29], all potentially serving Clinton's political aims at the time.

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 have been found to produce speech with greater amplitude modulations when instructed to produce clear speech [38] or when talking in noise (Lombard speech; [39]), presumably for reasons of increased intelligibility. As such, Clinton's speaking style during the three debates examined here may be the result of her extensive experience with public speaking.

One may also speculate about the absence of amplitude modulations in Trump's speech. Tian's recent analysis [40] 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 [40]; 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.

Despite the beneficial effects of rhythmic amplitude modulations on speech comprehension and speaker impressions, Clinton lost the American elections of 2016. Clearly, *how* you speak is not the only determinant of political success; *what* you say also contributes to political speakers' persuasiveness. Nevertheless, the present study has shown that speakers may (and do) use rhythmic amplitude modulations with the potential aim to enhance intelligibility and induce greater 'liking' and perceived charisma. This opens an opportunity for experimental studies targeting the perceptual effects of synthetically enhanced amplitude modulations on listeners' opinions about voice attractiveness, for instance.

4. Conclusions

This study shed light on the use of amplitude modulations 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.

5. Acknowledgements

This research was supported by a Gravitation grant from the Dutch Government to the Language in Interaction Consortium. Thanks goes to Robert Oostenveld for useful discussion about analysis procedures, and YouTube for making the digital recordings of the presidential debates available.

6. References

- [1] F. Pellegrino, C. Coupé, and E. Marsico, "Across-language perspective on speech information rate," *Language*, vol. 87, pp. 539-558, 2011.
- [2] O. Ghitza, "Behavioral evidence for the role of cortical θ oscillations in determining auditory channel capacity for speech," *Frontiers in psychology*, vol. 5, p. 652, 2014.
- [3] E. Jacewicz, R. A. Fox, and L. Wei, "Between-speaker and within-speaker variation in speech tempo of American English," *The Journal of the Acoustical Society of America*, vol. 128, pp. 839-850, 2010.
- [4] H. Quené, "Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo," *The Journal of the Acoustical Society of America*, vol. 123, pp. 1104-1113, 2008.
- [5] J. Verhoeven, G. De Pauw, and H. Kloots, "Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands," *Language and Speech*, vol. 47, pp. 297-308, 2004.
- [6] O. Ghitza and S. Greenberg, "On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence," *Phonetica*, vol. 66, pp. 113-126, 2009.
- [7] S. Greenberg and T. Arai, "What are the essential cues for understanding spoken language?," *IEICE Transactions on Information and Systems*, vol. E87-D, pp. 1059-1070, 2004.
- [8] S. Greenberg, "Speaking in shorthand – A syllable-centric perspective for understanding pronunciation variation," *Speech Communication*, vol. 29, pp. 159-176, 1999.
- [9] N. Ding, A. Patel, L. Chen, H. Butler, C. Luo, and D. Poeppel, "Temporal modulations in speech and music," *Neuroscience and Biobehavioral Reviews*, Online version, 2017.
- [10] O. Ghitza, "Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm," *Frontiers in Psychology*, vol. 2, 2011.
- [11] A.-L. Giraud and D. Poeppel, "Cortical oscillations and speech processing: emerging computational principles and operations," *Nature Neuroscience*, vol. 15, pp. 511-517, 2012.
- [12] J. E. Peelle and M. H. Davis, "Neural oscillations carry speech rhythm through to comprehension," *Frontiers in psychology*, vol. 3, 2012.
- [13] A. R. Bradlow, G. M. Torretta, and D. B. Pisoni, "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic

- talker characteristics," *Speech communication*, vol. 20, pp. 255-272, 1996.
- [14] O. Ghitza, "On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum," *Frontiers in psychology*, vol. 3, p. 238, 2012.
- [15] K. B. Doelling, L. H. Arnal, O. Ghitza, and D. Poeppel, "Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing," *Neuroimage*, vol. 85, pp. 761-768, 2014.
- [16] K. Aikawa and K. Ishizuka, "Noise-robust speech recognition using a new spectral estimation method "PHASOR"," in *Proceedings of Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, ed: IEEE, 2002, pp. 1-397-1-400.
- [17] A. Cutler, "Segmentation problems, rhythmic solutions," *Lingua*, vol. 92, pp. 81-104, 1994.
- [18] A. Cutler and D. Norris, "The role of strong syllables in segmentation for lexical access," *Journal of Experimental Psychology: Human perception and performance*, vol. 14, p. 113, 1988.
- [19] A. Cutler and S. Butterfield, "Rhythmic cues to speech segmentation: Evidence from juncture misperception," *Journal of Memory and Language*, vol. 31, pp. 218-236, 1992.
- [20] H. R. Bosker, "Accounting for rate-dependent category boundary shifts in speech perception," *Attention, Perception & Psychophysics*, vol. 79, pp. 333-343, 2017.
- [21] H. Quené and R. Port, "Effects of timing regularity and metrical expectancy on spoken-word perception," *Phonetica*, vol. 62, pp. 1-13, 2005.
- [22] M. P. Roncaglia-Denissen, M. Schmidt-Kassow, and S. A. Kotz, "Speech rhythm facilitates syntactic ambiguity resolution: ERP evidence," *PLoS One*, vol. 8, p. e56000, 2013.
- [23] K. Rothermich, M. Schmidt-Kassow, and S. A. Kotz, "Rhythm's gonna get you: regular meter facilitates semantic sentence processing," *Neuropsychologia*, vol. 50, pp. 232-244, 2012.
- [24] P. J. Essens and D.-J. Povel, "Metrical and nonmetrical representations of temporal patterns," *Perception & Psychophysics*, vol. 37, pp. 1-7, 1985.
- [25] C. Obermeier, S. A. Kotz, S. Jessen, T. Raettig, M. von Koppenfels, and W. Menninghaus, "Aesthetic appreciation of poetry correlates with ease of processing in event-related potentials," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 16, pp. 362-373, 2016.
- [26] C. Obermeier, W. Menninghaus, M. von Koppenfels, T. Raettig, M. Schmidt-Kassow, S. Otterbein, *et al.*, "Aesthetic and emotional effects of meter and rhyme in poetry," *Frontiers in Psychology*, vol. 4, 2013.
- [27] A. Rosenberg and J. Hirschberg, "Charisma perception from text and speech," *Speech Communication*, vol. 51, pp. 640-655, 2009.
- [28] O. Niebuhr, J. Voße, and A. Brem, "What makes a charismatic speaker? A computer-based acoustic-prosodic analysis of Steve Jobs tone of voice," *Computers in Human Behavior*, vol. 64, pp. 366-382, 2016.
- [29] F. D'Errico, R. Signorello, D. Demolin, and I. Poggi, "The perception of charisma from voice: A cross-cultural study," in *Proceedings of Affective Computing and Intelligent Interaction (ACII)*, ed: IEEE, 2013, pp. 552-557.
- [30] F. Ramus, M. Nespors, and J. Mehler, "Correlates of linguistic rhythm in the speech signal," *Cognition*, vol. 73, pp. 265-292, 1999.
- [31] E. Grabe and E. L. Low, "Durational variability in speech and the rhythm class hypothesis," *Laboratory Phonology*, vol. 7, 2002.
- [32] L. White and S. L. Mattys, "Calibrating rhythm: First language and second language studies," *Journal of Phonetics*, vol. 35, pp. 501-522, 2007.
- [33] V. Dellwo, "Rhythm and speech rate: A variation coefficient for ΔC ," in *Language and Language-Processing*, P. Karnowski and I. Sziget, Eds., Frankfurt am Main: Peter Lang, 2006, pp. 231-241.
- [34] A. Loukina, G. Kochanski, B. Rosner, E. Keane, and C. Shih, "Rhythm measures and dimensions of durational variation in speech," *The Journal of the Acoustical Society of America*, vol. 129, pp. 3258-3270, 2011.
- [35] (2016, Sept 28, 2016). *The First Presidential Debate: Hillary Clinton And Donald Trump (Full Debate) | NBC News*. Available: <https://www.youtube.com/watch?v=855Am6ovK7s>
- [36] (2016, Oct 9, 2016). *FULL VIDEO: Donald Trump vs Hillary Clinton - 2nd Presidential Debate*. Available: https://www.youtube.com/watch?v=h-gkBUBU_F4
- [37] (2016, Oct 19, 2016). *FULL VIDEO: Donald Trump vs Hillary Clinton - 3rd Presidential Debate*. Available: <https://www.youtube.com/watch?v=LsA6Gj8y8rU>
- [38] J. C. Krause and L. D. Braida, "Acoustic properties of naturally produced clear speech at normal speaking rates," *The Journal of the Acoustical Society of America*, vol. 115, pp. 362-378, 2004.
- [39] H. R. Bosker and M. Cooke, "Rhythm in plain and Lombard speech," presented at the 9th Speech in Noise Workshop, Oldenburg, Germany, 2017.
- [40] Y. Tian, "Disfluencies in Trump and Clinton first presidential debate," in *Proceedings of the conference Fluency and Disfluency Across Languages and Language Varieties*, Louvain-la-Neuve, 2017, pp. 106-109.