Then, what is charisma?
The role of audio-visual prosody in L1 and L2 political speeches

Hiroyuki Tanaka[1], Tamara Rathcke[2]

[1]Kotobaax, The Hague, Netherlands
[2]English Language and Linguistics, University of Kent, UK

kotobaax.hiro@gmail.com, t.v.rathcke@kent.ac.uk

Abstract

Charisma plays a significant role in political speeches, and determines the ability of a politician to carry an audience. While acoustic features of charisma have received some empirical attention, the contribution of visual prosody has been mostly neglected in studies focusing on features of a charismatic appearance. Unknown are also the audio-visual cues to charisma in non-native speakers. This small-scale study investigated speeches delivered by Donald Trump (L1 American English) and Arnold Schwarzenegger (L1 Austrian German, L2 American English). Video and audio recordings of their political speeches (around 25 min per speaker) and the transcripts were used. The use of pitch range, speech rate, emphatic stress and hand gestures was analysed. In order to establish the core means of the speakers' persuasive influence on their audiences, within-speaker comparisons were conducted for phrases with and without cheering from the audiences. The results showed some differences in the use of the audio-visual prosodic features between the L1 and L2 speaker as well as some similarities, and suggest that charisma is not easily attributable to a fixed set of prosodic means but may be best understood as a skillful modulation of audio-visual prosody in social interaction.

Index Terms: political speech, charisma, speech prosody, visual prosody, beat gesture, rhythm

1. Introduction

The term "charisma" originates from the Greek χάρισμα, meaning "a gift". The term has been traditionally used to refer to the ability of some persons to exert a strong influence on others, to make them believe in high personal competency and extraordinary powers of the speakers ([1]); these speakers are able to attract and retain large audiences ([2]). Although charisma is difficult to define precisely, listeners usually find it easy to identify if a speaker is charismatic or not.

One of the ways to observe if an audience regards a speaker as charismatic or not is to study the speaker's persuasiveness, and the link between persuasion and charisma has been previously discussed in the literature ([3]). Persuasiveness of a speaker plays a particular role in political speeches that might determine the rise or fall of a political party, make audiences take up required actions.

Although multimodality of charismatic appearance has been previously noted ([4]), most studies to date have concentrated exclusively on acoustic-prosodic features ([4, 5]). The main goal of the present study was to combine the auditory and the visual channels of charisma in political speeches, and to estimate the relative contribution of the two by taking into account an appreciative audience response such as cheering, applause and whistling. The second research question concerned the expression of charisma in non-native, as compared to native, speech. A non-native accent was expected to have its imprint in prosody ([6]), and to impact upon the use of prosodic means to charisma in contrast to non-verbal cues which were expected to be comparably used in both native and non-native politicians.

2. Method

2.1. Choice of speakers

The speakers selected for the present study were Donald Trump (L1 American English; hereafter DT) and Arnold Schwarzenegger (L1 Austrian German, L2 American English; hereafter AS). Both speakers are known to be popular and well-established public figures of American political scene, who have given various speeches for their electoral campaigns (presidential or governmental). Both can be described as charismatic as far as the scope of their persuasive popularity is concerned.

2.2. Speeches and sampling

Three recordings were selected, resulting in approximately 25 min material per speaker. Two shorter speeches were delivered by Trump (DT-1: a victory speech to his audience in Nevada after becoming the Republican nominee and DT-2: an electoral campaign speech on his economic policy in Pennsylvania). One longer speech was delivered by Schwarzenegger in the Republican convention in 2004 in his role as the governor of California. The videos stem from YouTube.

Appreciative audience responses during the speeches (such as applause, whistling, screaming) were identified and marked in time (negative reactions such as booing were not included, cf. [7]). Such audience responses are often considered a significant indicator of speaker persuasiveness ([7]). Syntactically complete phrases overlaid with cheers constituted the group of target phrases of this study (each phrase duration was between 1.2 and 6.8 sec). Control phrases comprised of several consecutive phrases with the overall duration of more than 6.8 sec, and did not involve any simultaneous cheering or disapproving noises. Four sets of control phrases were chosen from DT-1 and DT-2 recordings (with the total duration of 190 sec) and three sets were selected from AS (with the total duration of 125 sec; for more detail on this procedure see [8]).

2.3. Prosodic features

Previous research has established that an increased pitch range, moderately fast speaking rate, and emphatic stress are the core features of prosodic importance for the creation of charisma, at least in American English – the variety under investigation in the present study ([2,5]). Accordingly, these three prosodic features were isolated in target and control phrases, analysed and transcribed using Praat.

2.4. Annotation of gestures

This study concentrated on speech-accompanying hand gestures only. A gesture was defined as a movement or series of continuous movements of one or two hands simultaneously during speech. Table 1 summarises previous accounts of such gestures, in comparison to the account taken up in the present study. Occasional cases of doubts where a hand movement could be interpreted as two different gestures, both gesture types were annotated for the same phrase. Importantly, the duration of gesturing during speech was not measured, only the gesture types in target vs. control phrases were annotated.

Table 1. *A summary of gestural classifications in previous accounts (account-1 ([9,10]); account-2 ([11]); account-3 ([12]) and the present study.*

Account-1	Account-2	Account-3	Present annotation
Baton-like/ Ideographic	-	Beats	Rhythmic (**RG**)
Deictic	Demonstra-tive	Deictic	Indexical (**IG**)
Symbolic/ Emblematic	Connotative	Metaphoric	Connotative (**CG**)
Illustrating/ Icono-graphic	Mimic/ Symbolic	Iconic	Denotative (**DG**)

3. Results and discussion

3.1. Speaker comparison

Table 2 summarises and compares values of the relevant prosodic features, Figure 2 displays the overall distribution of the four gesture types measured for the two speakers. Overall, prosodic measurements support our expectation and show a discrepancy primarily in the pitch range between the two speakers: Although the physiological pitch range of the two speakers is comparable (ranging between 75 and 380 Hz), the functional pitch range is substantially smaller in the L2-speaker AS (due to his more reluctant use of high pitch), which is in line with the findings on L1-German speakers of English ([6]). However, in contrast to our expectation, the two speakers also showed substantially different personal preferences for their speech accompanying gestures, with DT having a strong preference for rhythmic gestures and AS preferring mainly the indexical gestures. Moreover, DT produced emphatic stresses much more frequently than AS.

3.2. Comparison of target vs. control phrases

Table 3 shows the distribution of prosodic features across target vs. control phrases of the two speakers. Emphatic stress occurred exclusively in the target phrases (i.e. accompanied by cheers from the audience) in case of AS; similarly, only one emphatic stress was observed outside target phrases in DT. It was likely to be carried by pitch accents with an extended pitch range. Target phrases also displayed a slight increase of speech rate. In sum, prosodic features previously identified as cueing charismatic speech ([2,5]) were observed in target rather than in control phrases.

Table 2. *Speech rate and frequency for the two speakers.*

Measurement	DT	AS
Mean speech rate	4.0 syll/sec	4.0 syll/sec
Min F0 produced	74 Hz	75 Hz
Mean min F0	106 Hz	115 Hz
Max F0 produced	387 Hz	376 Hz
Mean max F0	297 Hz	241 Hz
Mean pitch range	18 st	13 st
Emphatic stress use	3.92 times/min	2.59 times/min

Table 3. *Percentage of emphatic stress occurrences and mean values for speech rate (syll/sec) and pitch range (st) measured in target vs. control phrases of DT vs. AS.*

Measurement	DT		AS	
	target	control	target	control
Emphatic stress	99 %	1 %	100 %	0 %
Speech rate	4.2	3.8	4.1	3.9
Pitch range	28	18	24	13

Figure 1: *Distribution of the gesture types across all phrases produced by DT and AS.*
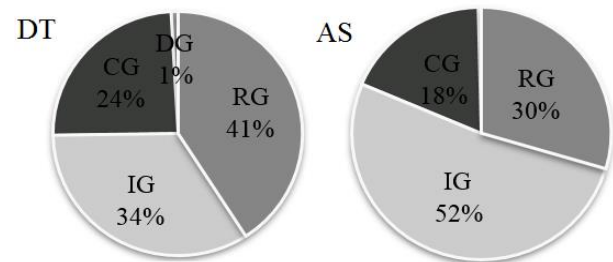
4. Conclusions

These results further suggest that charisma is not easily attributable to a fixed set of prosodic means but may be best understood as a skillful modulation of audio-visual prosodic means in social interaction. More fine-grained analyses of gesture timing, inclusion of disapproving responses, cross-cultural comparisons and a larger database of political speeches will help to shed brighter light on how charisma arises in different socio-cultural contexts.

References

[1] Awamleh, R. & Gardner, W. L. (1999). Perceptions of Leader Charisma and Effectiveness: The Effects of Vision Content, Delivery, and Organisational Performance. Leadership Quarterly, 10(3), 345-373.

[2] Rosenberg, A. & Hirschberg, J. (2009). Charisma Perception from Text and Speech. *Speech Communication*, 51, 640-655.

[3] Dewan, T., Humphreys, M., & Rubenson, D. (2013). The Elements of Political Persuasion: Content, Charisma and Cue. The Economic Journal, 124, 257-292.

[4] Signorello, R., D'Errico, F., Poggi, I., Demolin, D., & Mairano, P. (2012). Charisma Perception in Political Speech: A Case Study. In: Mello, H., Pettorino, M., & Raso, T., eds. Proceedings of the 7th GSCP International Conference: Speech and Corpora, February 2012. Firenze University Press, pp. 343-348.

[5] Niebuhr, O., Brem, A., Novák-Tót, E., & Voβe, J. (2016). Charisma in Business Speeches: A Contrastive Acoustic-Prosodic

Analysis of Steve Jobs and Mark Zuckerberg. In: *Proceedings of the 8th International Conference of Speech Prosody, Boston*.

[6] Mennen, I., Schaeffler, F., Docherty, G. (2012). Cross-language differences in fundamental frequency range: A comparison of English and German. Journal of Acoustical Society of America 131(3), 2249.

[7] Guerini, M., Strapparava, C., & Stock, O. (2008). CORPS: A Corpus of Tagged Political Speeches for Persuasive Communication Processing. Journal of Information Technology & Politics, 5(1), 19-32.

[8] Tanaka, H. (2016). *Acoustic and Gestural Analyses of Political Speeches for Effective Public Speaking Training in TESOL.* Unpublished MA thesis, University of Kent.

[9] Ekmen, P. & Friesen, W. (1969). The Repertoire of Nonverbal Behavior: Categories, Origins, Usage and Coding. Semiotica, 1, 49-98.

[10] Efron, D. (1972). Gesture, Race and Culture. The Hague: Mouton & Co.

[11] Wundt, W. (1973). The Language of Gestures. The Hague: Mouton & Co.

[12] McNeill, D. (1992). Hand and Mind. Chicago: University of Chicago Press.