Culture-specific differences in multimodal aspects of political speeches

Maciej Karpiński[1], Ewa Jarmołowicz-Nowikow[1], Wojciech Laskowski[2], Katarzyna Klessa[1]

[1] Adam Mickiewicz University, Poznań,

[2] WSB Merito University, Poznań

{maciej.karpinski | ewa.jarmolowicz-nowikow | katarzyna.klessa}@amu.edu.pl,
wojciech.laskowski@wsb.poznan.pl

Political speeches are willingly explored by linguists not only because of their importance to the public and social life and potential influences on economy, but also because of the speakers themselves, sometimes unique personalities, showing extraordinary rhetorical skills and high persuasive abilities. However, one may notice certain culture- and language-related differences in the behaviour of speakers. This pertains not only to the way they speak, but obviously to their gesturing patterns. Although many universal tendencies can be observed in the prosodic expression of affect (Scherrer et al. 2005), some studies (Chen 2005; Thompson & Balkwill 2006; Karpiński et al. 2022) as well as our preliminary scrutiny of the material show that, in the domain of paralinguistic prosody, pitch usage in may be culture-specific, while in the gestural domain, e.g., gesture amplitude and category may be culture-dependent (Kita 2009). In principle, one may assume that higher pitch and larger gestures correlate with higher affective engagement, but the detailed usage of these expressive dimensions by speakers may differ, depending on a number of factors, including not only "political culture" but also age or personality.

In the present pilot contribution, we analyse political debates at the final stages of presidential elections in the USA (2020) and Poland (2015). Both the American and Polish debates were moderated by TV broadcast moderators who asked questions and controlled the flow of the debate. The total duration of the US American debate, including introduction and ending was 1h 33 min, while the Polish one took 1h 29min. The recordings were obtained from public social media (the official YouTube channels of broadcasting media) as freely available material. They were manually annotated in ELAN (Wittenburg et al., 2006) for the usage of gestures produced with index finger extended including their categories (based on their shape and manner of production) and selected detailed features. Annotation was carried out and cross-checked by two experts. The total of 625 pointing gestures were annotated in all the candidates (177 and 49 in the Polish ones, 210 and 189 in the American ones). The distributions of the gesture categories used by the American and Polish candidates clearly differ but, in both cases, they are dominated by jabbing gestures – movements along the line of the straightened index finger in such a way as if the speaker wanted to push someone or something hard and quickly with the tip of the finger (422 occurrences). Certain categories are unique only to one speaker.

In order to find whether the usage of certain gesture categories or gestural features differ in terms of pitch in co-occurring portions of speech, we used Praat (Boersma & Weenink, 1992-2021) to automatically extract all the portions of signal. The extracted signals were scrutinised for potential quality issues, including speech overlaps or speaker being not heard (or not speaking) within the time limits of a given gesture. As a result, 73 of them were excluded from further analyses. Another Praat script was used to analyse pitch and related parameters, using standard Praat autocorrelation algorithm. The maximum, minimum, and mean *f0* values as well as standard deviation, slope and relative pitch range, were calculated the extracted stretches of speech.

The dominating category of "jabbing" gestures (a subcategory of deictic gestures, distinguished for the purpose of this study) (1) was juxtaposed against the remaining major categories of gestures produced with extended index finger (2). For mean *f0* values, linear mixed models indicate a significant effect of gesture category for US English. The mean *f0* was found to be significantly higher for the category 1, while for Polish no significant difference was found in this respect (Anova, F=6.3435, p=0.012; cf. Figure 1). For the maximum *f0* values, significant interaction was shown between language and gesture category (p = 0.03), and a significant effect of gesture category (p = 0.02), while no main effect of language was found (p > 0.5) (cf. Fig. 2). In terms of the interactions between *f0* and gesture amplitude, the main effect for gesture amplitude was observed with p < 0.001. In Figure 2, the values of the maximum *f0* predicted using linear mixed models are displayed as related to gesture amplitude (for each successive level of amplitude a respective rise of *f0* maximum values). Gesture amplitude was defined as the number of the sections of gesture space involved in the realisation of a given gesture.

Our pilot results show some culture-specific features in the speaking styles of presidential candidates both in terms of pitch and gesture usage. In further research, we plan to pay more attention to the detailed shapes of pitch contours as well as to pitch and duration-related parameters in the gesture-preceding portions of speech. This approach and techniques will be further used in the studies on the parliamentary speeches, with a higher number of speakers as well as with different languages.

Index Terms: political debates, affective prosody, gestures produced with index finger extended

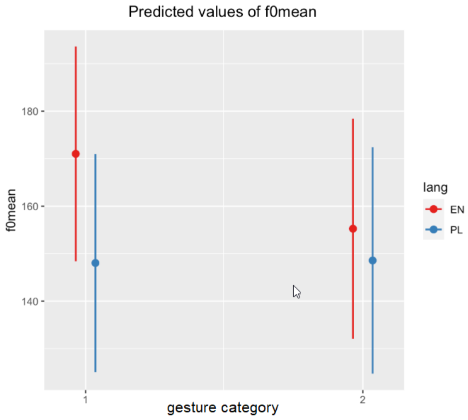

Figure 1. *The predicted values of f0 for jabbing gestures (1) and the remaining gesture produced with index finger extended (2) by American English and Polish speakers.*

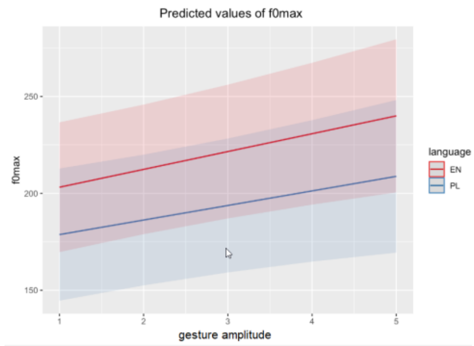

Figure 2: *The predicted values of f0 max depending on gesture amplitude for Polish and American English speakers*

Acknowledgement

MuMo Stance project is funded by the Polish National Science Foundation (NCN 2018/31/G/HS2/03633) and Deutsche Forschunggemeinschaft. We are grateful to Bettina Braun for her advice on data analysis.

References

Boersma, P., & Weenink, D. 1992–2021. Praat: doing phonetics by computer [Computer program]. Ver. 6.2. Retrieved on 15 November 2021 from: https://www.praat.org

Chen, A. J. 2005. *On the universal and language-specific perception of paralinguistic intonational meaning*. Utrecht: LOT.

Karpiński, M., Jarmołowicz-Nowikow, E., Klessa, K. 2022. High-pitched prominences in the speeches of male Polish members of parliament. Proc. Speech Prosody 2022, doi: 10.21437/SpeechProsody.2022-75, 367-371.

Kita, S. 2009. Cross-cultural variation of speech-accompanying gesture: A review. Language and cognitive processes, 24(2), 145-167.

Müller, C., Bressem, J., & Ladewig, S. H. 2013. 45. Towards a grammar of gestures: A form-based view. In: Body – Language – Communication, vol. 1. De Gruyter Mouton, 707-733.

Scherer, K. R., Banse, R. and Wallbott, H. 2001. Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, 32, 76-92.

Thompson, W. F., & Balkwill, L. L. 2006. Decoding speech prosody in five languages. *Semiotica*, 158, 407–424.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. 2006. ELAN: a professional framework for multimodality research. In *Proceedings of the 5th Language Resources and Evaluation Conference*, Genoa, Italy, 1556-1559.