THE PROSODY OF HIGH AND LOW AFFECTIVE ENGAGEMENT IN POLISH AND GERMAN PARLIAMENTARY SPEECHES

Maciej Karpiński, Katarzyna Klessa, Brygida Sawicka-Stępińska, Hanna Kasperek

Adam Mickiewicz University, Poznań
{maciej.karpinski | katarzyna.klessa | brygida.sawicka-stepinska | hanna.kasperek}@amu.edu.pl

ABSTRACT

Parliamentary speeches tend to stick to predetermined topics and structures but they are often characterised by a high degree of emotionality and expressivity. Meant to sound convincing and reliable, they still raise emotions in the audience. In the era of international parliamentarism, it is important to understand the role of (para)linguistic and cultural settings in the perception of such speeches. This study is based on a multimodal corpus of parliamentary speeches from the German Bundestag and Polish Sejm, including multilayer annotations of speech and gestures. We inspect prosodic features of specific chunks of the recordings marked as speakers' "high" or "low" affective engagement areas. For the above engagement areas, the variability of pitch and timing features of the utterances is analysed. Our results show that despite progressing internationalisation and collaboration of politicians, certain differences in the paralinguistic prosody of German and Polish speeches remain significant.

Keywords: paralinguistic prosody, affective engagement, Polish, German, parliamentary speech

1. INTRODUCTION

Parliamentary debates have the potential to influence the direction of political and socio-economic changes. While the content of the speeches is of obvious importance, the way they are delivered may be at least equally significant [1]. Their persuasive power lies not only in the selection of words and in the argument structure but also in their prosodic characteristics [2, 3, 4, 5], with a substantial contribution from attitudinal, affective prosody [6, 7, 8]. Prosodic symptoms of emotional arousal typically occur in significant, ideologically engaged debates. However, speakers definitely differ in their abilities to control and hide emotions, as well as in the very strategy of showing or not showing them. Some act and fake emotions more consciously, while others are honestly engaged. Finally, their default, everyday individual speaking styles may differ, being perceived as generally more or less emotional. Therefore, the analysis of affective engagement may not only be important for correlating it with, e.g., persuasive power, but also helpful in decoding speakers' intentions and judging their credibility. In the present study, selected differences in the prosodic characteristics of Polish and German parliamentary speeches are explored. The areas of high and low affective engagement are marked manually on the basis of audio-visual scrutiny by Polish and German expert teams [9, 10]. Selected pitch- and duration-related parameters are measured for the Polish and German speakers in the areas of high and low affective engagement and hypothesized as cues for identifying these areas. We explore them using lineaer mixed effects regression models with language and affective engagement as fixed factors, and speaker as random effects. The results are discussed in the context of public political communication and multimodal analyses.

2. STUDY MATERIAL

The material under study comes from a multimodal corpus of German and Polish parliamentary speeches, developed within the MuMo Stance project (see Acknowledgement). The corpus comprises audio and video recordings of speeches delivered by the members of parliament during the 2020 budgetary debates, along with multi-tier, time-aligned annotation and segmentation on the level of the phrase, word, syllable, and phone. For speech transcription and segmentation, orthographic transcripts of parliamentary speeches available in the archives of the Polish Sejm and German Bundestag were used. They were time-aligned to inter-pausal units and manually adjusted by experienced annotators as they often lack phonetically important details (like repetitions or information on fillers). The inter-pausal units were used as the input for automatic segmentation into words, syllables, and phones. The segmentation of the Polish material was performed using the ANNPRO desktop module of CLARIN-

PL tools [11], while the German material was segmented using WebMaus [12]. In both cases, manual corrections of segment boundary positions were introduced based on visual inspection of spectrograms, auditory inspection and guidelines specified, for example, in [13]. Speakers' behaviour is described based on perception judgements, following the expressive movement framework approach [14]. The framework involves annotation of audio-video recorded utterances in terms of the areas of speakers' high and low affective engagement (HAE and LAE, respectively). Such areas are manually annotated using ELAN [15] in a three-fold procedure. First, each speech is divided into interactive units, i.e., sequences of speech interrupted either by applause or by any audible or visible interjections which receive a reaction from the speaker. Then, within each unit, HAE and LAE areas are distinguished on the basis of the perceived behaviour expressive intensity [16, 17]. As speakers vary in their ways and strategies of emotional expression, the annotators must gain an understanding of a speaker's unique style, as well as their personal benchmark of expression. For this reason, affective engagement areas were determined individually for each speaker, assessed according to their overall performance and forms of expression presented during the speech. The greater the expressive resources that are mobilized, the greater the level of affective engagement, and thus such passages are recognised as HAE areas; the remaining parts are described as LAE areas. The third step of the expressive movement description consists of tandem sessions, during which final version of each annotation is calibrated. Should discrepancies occur, they are discussed during a meeting of the annotators involved. The entire corpus was inspected for the occurrence of HAE and LAE areas. For the present study, only the speeches of male speakers containing both HAE and LAE areas were selected from the budgetary debate material. As a result, seven German and nine Polish speeches meeting the above criteria have been used. Polish speeches tend to be shorter (from 67s to 1567s, mean=454, sd=480 for Polish and from 236s to 977s, mean=575, sd=291 for German), while the proportion of LAE and HAE is almost equal (16:84 for Polish and 17:83 for German, respectively).

3. METHODS AND ANALYSES

In the present study, we analyse how selected prosodic characteristics of HAE and LAE areas differ between Polish and German parliamentary speeches. The following prosodic parameters were taken into account as potentially influenced by the affective state of the speaker: mean pitch frequency, pitch frequency standard deviation [18, 19], mean intensity, intensity standard deviation [20, 21], vocalic, consonantal, and syllabic nPVI [e.g., [22, 23, 24]], and TGA duration slope and intercept linear regression values [25]. For quantitative analyses, all the HAE and LAE areas were sampled using a 5 second time window which gave the total of 654 samples (537 from LAE and 117 from HAE areas) for German and 505 (424 from LAE and 81 from HAE areas) for Polish. For each sample, the prosodic parameters were extracted or calculated using Annotation Pro plugins [26]. Pitch frequency values were extracted from pitch-smoothed [27] signals using Praat's [28] standard autocorrelation method with the frequency range from 75Hz to 350Hz. The upper limit was selected on the basis of previous studies of pitch in male parliamentary speakers e.g., [29]. A custom-made plugin was used to extract mean pitch and standard deviation values from the samples. Intensity values were extracted using standard Praat algorithm with the mean energy averaging method, and imported into Annotation Pro. However, they turned out to systematically differ between the languages in a way that may have resulted from some differences between recording setups in the two parliaments. While potentially useful for other purposes, here they were excluded from further analyses. The TGA slope and intercept values were calculated with the Annotation Pro+TGA plugin [30] for the entire affective engagement areas and then inherited by the samples taken from those areas. Table 1 provides basic descriptive statistics for the variables under study. Linear mixed effects regression models (lmar in R) were fitted for each variable with language and affective engagement, their interactions as fixed factors and speaker as random effects (random intercept and slope for affective engagement), cf. [31]. P-values were adjusted using the Benjamin-Hochberg correction [32]. The results of analyses with $p<0.05$ are listed in Table 2. In spite of their apparent significance, most of the main effects may be disputable as they may result from systematic differences between the two languages or from strong variation among individual speakers, as their number is relatively limited. However, the main interactions of the mean pitch value, mean syllabic nPVI, and mean vocalic nPVI interactions, may provide an insight into some relevant phenomena. The remaining factors did not have a significant effect ($p>0.07$). As shown in Fig. 1., the distribution

of the mean pitch values in samples from LAE and HAE areas differs between the languages, and a clear difference can be spotted between LAE and HAE values for the German language. In Fig. 2, the distribution of the syllabic nPVI is shown, and Fig. 3 illustrates the data on the vocalic nPVI from the LAE and HAE in German and Polish speakers. The distribution of the syllabic nPVI values within languages does not differ much between LAE and HAE but there is a significant difference between the languages. The same applies to the vocalic nPVI where the distributions are relatively similar within languages but differ between them, with lower means for Polish and more dispersion of the values for the LAE samples. It may be partially due to the phonological duration present in German but absent in Polish. Nevertheless, at this stage, one cannot exclude the impact resulting from factors related to different traditions of public speaking in Germany and in Poland. Higher TGA mean slope values are observed for Polish than in German. For both languages, the slope is lower in HAE which means less deceleration in interpausal time groups than in LAE. Basic descriptive statistics of the explored parameters in the material under study are listed in Table 1.

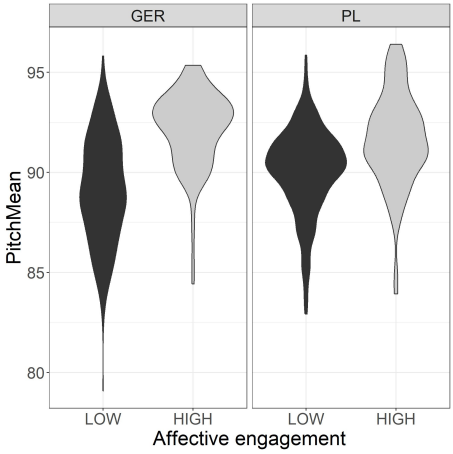

Figure 1: Mean pitch values in the LAE and HAE in German and Polish parliamentary speeches.

4. CONCLUSIONS

Political speeches are a peculiar instance of language use where speakers address a very wide, often physically absent audience. They tend to be well-controlled and are often pre-prepared but still emotional. However, the expression of emotions may be influenced or channeled by the social and cultural factors, including, e.g., the parliamentary

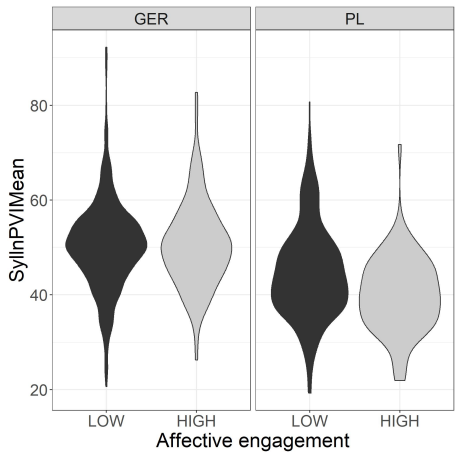

Figure 2: Mean syllabic nPVI values in the LAE and HAE in German and Polish parliamentary speeches.

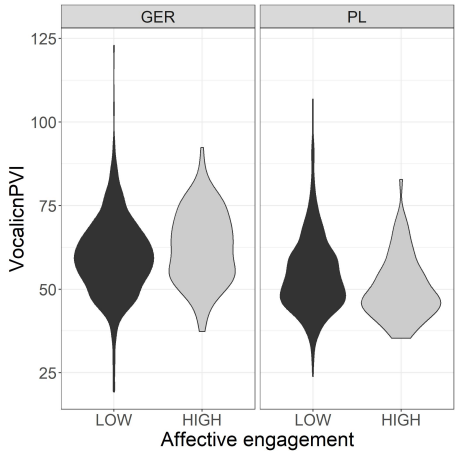

Figure 3: Mean vocalic nPVI values in the LAE and HAE in German and Polish parliamentary speeches.

rules of conduct. Therefore, because of its peculiar characteristics and contextual factors, this variety still requires exploration in spite of the huge body of already available studies on emotional speech prosody. In the present paper, some differences are shown in how the states of high and low affective engagement are reflected in the prosody of German and Polish parliamentary speakers. As expected, the language itself, along with its culture-related, paralinguistic component, turns out to be a very strong factor. While the meaning of the main effects may be disputable, it may be clarified with a larger group of participants. The main interactions of the mean pitch value, mean syllabic nPVI, and mean vocalic nPVI, seem to point to a promising direction for further research on both

		GERMAN		POLISH	
	Affective engagement	Mean	Standard deviation	Mean	Standard deviation
Vocalic nPVI	LAE	59.80	11.47	50.33	11.27
	HAE	62.67	10.80	54.41	9.38l
Consonantal nPVI	LAE	45.91	10.52	48.43	8.56
	HAE	44.88	11.57	48.53	7.45
Syllabic nPVI	LAE	50.34	9.29	44.67	9.74
	HAE	50.52	9.11	40.04	8.00
Mean Pitch (ST)	LAE	89.15	2.62	90.09	2.12
	HAE	92.17	1.90	91.64	2.36
SD of Pitch (ST)	LAE	2.42	0.94	2.68	0.87
	HAE	2.15	0.79	2.80	0.83
TGA Mean Slope	LAE	0.02	0.08	0.09	0.16
	HAE	0.00	0.05	0.05	0.10
TGA Mean Intercept	LAE	195.82	45.38	166.20	44.36
	HAE	221.28	44.20	167.69	36.14

Table 1: Basic descriptive statistics of the explored parameters in the material under study.

	analysis	p-value	adjusted p-value	DenDF	F
1	PitchMean_main_AffEng	0.000	0.000	15.70	85.92
2	PitchMean_interaction	0.04606	0.04606	15.70	4.69
3	SlopeMean_main_AffEng	0.04076	0.0439	8.57	5.79
4	SlopeMean_main_Language	0.00039	0.00109	14.16	21.30
5	InterceptMean_main_AffEng	0.0199	0.02567	12.9	7.05
6	InterceptMean_main_Language	9.00E-04	0.00193	13.71	17.76
7	SyllnPVIMean_main_AffEng	0.02017	0.02567	1091.89	5.41
8	SyllnPVIMean_main_Language	9.00E-05	0.00041	13.58	30.22
9	SyllnPVIMean_interaction	0.03534	0.04123	1091.89	4.44
10	VocalicnPVI_main_Language	0.00097	0.00193	11.12	19.72
11	VocalicnPVI_interaction	0.00857	0.01333	78.55	7.27

Table 2: The significance of interactions and main effects (only the cases of p<0.05 are listed). AffEng stands for Affective Engagement of the speaker, labelled as "low" or "high".

individual differences (including gender or political orientation) and more general tendencies regarding the roles of the prosodic parameters as cues to the identification of low and high affective engagement. Our research will include the gestural component of the speeches and facial expression as well [33]. Finally, a cross-cultural testing of the perception of the prosodic characteristics will be carried out [34].

5. ACKNOWLEDGEMENTS

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

6. REFERENCES

[1] I. Poggi, F. D'Errico, L. Vincze, and A. Vinciarelli, Multimodal Communication in Political Speech Shaping Minds and Social Action: Int. Workshop, Political Speech 2010. Springer, 2013, vol. 7688.

[2] P. Touati, "Prosodic Aspects of Political Rhetoric," in ESCA Workshop on Prosody, 1993.

[3] J. B. Hirschberg and A. Rosenberg, "Acoustic/Prosodic and Lexical Correlates of Charismatic Speech," Columbia Academic Commons, 2005.

[4] G. Kišiček, "Persuasive power of prosodic features," Argumentation and Advocacy, vol. 54, no. 4, pp. 345–350, 2018.

[5] J. Michalsky, H. Schoormann, and T. Schultze, "Towards the prosody of persuasion in competitive negotiation. The relationship between f0 and negotiation success in same sex sales tasks," in Interspeech, 2019, pp. 311–315.

[6] F. D'Errico, R. Signorello, D. Demolin, and

I. Poggi, "The Perception of Charisma from Voice: A Cross-Cultural Study," in Humaine Association Conference on Affective Computing and Intelligent Interaction. IEEE, 2013, pp. 552–557.
[7] R. L. Mitchell and E. D. Ross, "Attitudinal prosody: What we know and directions for future study," Neuroscience & Biobehavioral Reviews, vol. 37, no. 3, pp. 471–479, 2013.
[8] E. Strangert, "Prosody in public speech: analyses of a news announcement and a political interview," in 9th European Conference on Speech Communication and Technology, 2005.
[9] H. Kappelhoff and C. Müller, "Embodied meaning construction: Multimodal metaphor and expressive movement in speech, gesture, and feature film," Metaphor and the social world, vol. 1, no. 2, pp. 121–153, 2011.
[10] C. Müller and H. Kappelhoff, "Cinematic metaphor," in Cinematic Metaphor. De Gruyter, 2018.
[11] K. Klessa, D. Koržinek, B. Sawicka-Stępińska, and H. Kasperek, "ANNPRO: A Desktop Module for Automatic Segmentation and Transcription," in Language and Technology Conference. Springer, 2022, pp. 65–77.
[12] T. Kisler, U. Reichel, and F. Schiel, "Multilingual processing of speech via web services," Computer Speech & Language, vol. 45, pp. 326–347, 2017.
[13] P. Machač and R. Skarnitzl, Principles of Phonetic Segmentation. Epocha, 2009.
[14] D. Horst, F. Boll, C. Schmitt, and C. Müller, "Gesture as interactive expressive movement: Inter-affectivity in face-to-face communication," in Body - Language - Communication: An International Handbook on Multimodality in Human Interaction, vol. 38.2. De Gruyter Mouton, 2014, pp. 2112–2126.
[15] P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes, "Elan: A professional framework for multimodality research," in 5th LREC, 2006, pp. 1556–1559.
[16] C. Müller, Metaphors Dead and Alive, Sleeping and Waking: A Dynamic View. Univ. of Chicago Press, 2008.
[17] C. Müller and S. Tag, "The Dynamics of Metaphor: Foregrounding and Activating Metaphoricity in Conversational Interaction," Cognitive Semiotics, vol. 6, no. s1, pp. 85–120, 2010.
[18] J. Vroomen, R. Collier, and S. J. Mozziconacci, "Duration and intonation in emotional speech." in Eurospeech, 1993.
[19] B. J. Dietrich, M. Hayes, and D. Z. Oâbrien, "Pitch Perfect: [v]ocal Pitch and the Emotional Intensity of Congressional Speech," American Political Science Review, vol. 113, no. 4, pp. 941–962, 2019.
[20] I. R. Murray and J. L. Arnott, "Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion," JASA, vol. 93, no. 2, pp. 1097–1108, 1993.
[21] C. Pereira, "Dimensions of emotional meaning in speech," in ISCA ITRW Speech and Emotion, 2000.
[22] L. E. Low, E. Grabe, and F. Nolan, "Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English," Language and Speech, vol. 43, no. 4, pp. 377–401, 2000.
[23] L. Quinto, W. F. Thompson, and F. L. Keating, "Emotional Communication in Speech and Music: The Role of Melodic and Rhythmic Contrasts," Front. Psychol., vol. 4, p. 184, 2013.
[24] M. Goudbeek and M. Broersma, "Rhythm in vocal emotional expressions: the normalized pairwise variability index differentiates emotions across languages," in Bi-Annual Conference of ISRE, Geneva, 2015.
[25] D. Gibbon, "TGA: a web tool for Time Group Analysis," in Proc. of the TRASP Workshop, Aix-en-Provence, 2013, pp. 66–69.
[26] K. Klessa, M. Karpiński, and A. Wagner, "Annotation Pro - a new software tool for annotation of linguistic and paralinguistic features," in Proc. TRASP Workshop, Aix-en-Provence, 2013, pp. 51–54.
[27] "Praat Vocal Toolkit," https://https://www.praatvocaltoolkit.com, visited 5-Jan-23.
[28] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer (version 4.3.14) [computer program]," 2014, (visited 20-Dec-17). [Online]. Available: http://www.praat.org
[29] M. Karpiński, K. Klessa, and E. Jarmołowicz-Nowikow, "High-pitched prominences in the speeches of male Polish members of parliament," in Proc. of the 11th Speech Prosody, Lisbon, 2022, pp. 367–371.
[30] K. Klessa and D. Gibbon, "Annotation Pro+ TGA: automation of speech timing analysis," in Proc. of the 9th International LREC, 2014, pp. 1499–1505.
[31] R. H. Baayen, D. J. Davidson, and D. M. Bates, "Mixed-effects modeling with crossed random effects for subjects and items," Journal of memory and language, vol. 59, no. 4, pp. 390–412, 2008.
[32] D. Thissen, L. Steinberg, and D. Kuang, "Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons," Journal of Educational and Behavioral Statistics, vol. 27, no. 1, pp. 77–83, 2002.
[33] J. Streeck, "Gesture in Political Communication: A Case Study of the Democratic Presidential Candidates During the 2004 Primary Campaign," Research on Language and Social Interaction, vol. 41, no. 2, pp. 154–186, 2008.
[34] A. Cullen, A. Hines, and N. Harte, "Building a Database of Political Speech: Does Culture Matter in Charisma Annotations?" in Proc. of the 4th International Workshop on Audio/Visual Emotion Challenge, 2014, pp. 27–31.