When Gestures Enter the Game, Prosody Breaks the Rules

Aleksandra Ćwiek[1], *Šárka Kadavá*[1], *Katarzyna Stoltmann*[2], *Wim Pouw*[3], *Susanne Fuchs*[1]

[1] Leibniz-Centre General Linguistics, Berlin [2] adesso SE

[3] Donders Institute for Brain, Cognition and Behaviour, Nijmegen

{cwiek,kadava,fuchs}@leibniz-zas.de, wim.pouw@donders.ru.nl

Gestures and speech are intimately linked, despite operating at different time scales (Grimme et al., 2011). It has been demonstrated that gesture-speech coordination tends to occur at prominent syllables (Franich, 2022; Krivokapić et al., 2017; Rochet-Capellan et al., 2008) and that gestures enhance prominence, especially in languages with flexible lexical stress (Cutler, 2005). It is less clear how this gesture-speech interaction works in languages with fixed stress, such as Polish. Here we investigate whether imposed rhythmicity on both speech and gesture can override or modulate the prosodic requirements of Polish and lead to a shift in stress patterns.

We investigated the production of counting-out rhymes in Polish, which offer an opportunity to explore the relationship between gesture and speech, because they naturally involve pointing gestures and high rhythmicity. Polish, with its fixed primary stress on the penultimate syllable (Dłuska, 1976; Jassem, 1962), presents an interesting case for this investigation. While only a small percentage of exceptions exist in the lexicon (Peperkamp et al., 2010), recent research suggests that native speakers can produce stress shifts even onto initial positions (Osowicka-Kondratowicz, 2021, 2022), which are typically occupied by secondary stress in words of four or more syllables. However, the status and acoustic expression of secondary stress in Polish is still debated (Dogil & Williams, 1999; Malisz & Żygis, 2018; Rubach & Booij, 1985).

We aim to evaluate the flexibility of stress production in a manual rhythmic task. We recorded six counting-out rhymes from 11 native Polish speakers with and without pointing gestures while manipulating speech rate (normal vs. fast) and pointing hand (left vs. right) within the pointing condition (Stoltmann & Fuchs, 2017). Speech data were processed automatically using WebMAUS (Kisler et al., 2017), followed by manual correction for syllable intervals. For each syllable, we measured duration, F0 peak, and amplitude envelope peak – as potential stress markers in Polish (Ćwiek & Wagner, 2018; Dogil & Williams, 1999; Malisz & Żygis, 2018) – as dependent variables. The syllables were also annotated for predicted stress across prosodic words (primary vs. secondary vs. no).

We used Bayesian inference to model the data, with predicted stress as the predictor in all models (with primary stress as baseline), alongside random intercepts for participant and rhyme with maximal slopes. Three models were computed for each dependent variable: (1) comparing data with and without movement by including this condition as a binary fixed factor; (2) analyzing only data without pointing by adding an extra fixed predictor for rhyme familiarity; (3) examining only data including pointing, using additional fixed predictors of rhyme familiarity, speed, and pointing hand. Duration was log-scaled, F0 was used in raw values in Hz (with sex as a controlled fixed factor), and amplitude envelope was z-scored. We set weakly informative priors and ran all models with 4 chains á 8,000 iterations after warmup. The scripts and models are available in the OSF repository: https://osf.io/xcqu4

The findings on predicted stress in all models are inconclusive. Neither secondary nor no stress shows a significant difference from primary stress. This indicates that participants not only shift prominence to syllables with secondary stress, which would be phonologically acceptable, but mark previously unaccented syllables. This is consistent with the results by Osowicka-Kondratowicz (2021, 2022), demonstrating that native speakers tolerate even violations of three-syllabic words moving stress from the penultimate to initial positions, as also evidenced in one of the rhymes in our data (cf. Figure 1). Additionally, the 95% credible interval (CrI) for no stress is consistently skewed towards a lower value than the intercept. Meanwhile, in all models, the 95% CrI for secondary stress exhibits significant width. This suggests that the posterior secondary stress may be considerably higher or lower than the intercept; therefore, making it highly unpredictable based on the current data, model, and priors.

In models including speech with pointing, we observed an effect of speed condition on syllable duration (β = -0.07 [-0.08, -0.05]). Syllables are shorter in the faster condition. Furthermore, we found a subtle but reliable effect of hand on duration (β = 0.01 [0.00, 0.01]), indicating that syllables are longer when participants use their left hand to count out. Hand is also a reliable predictor for the amplitude envelope peak (β = -0.06 [-0.11, -0.01]), suggesting that envelope peaks tend to be higher when participants use their right hand to count out. These effects highlight the relationship between gesture and speech, as well as the adaptation that speech as a faster system undertakes to adjust to gesture as a slower system. During the conference, we will present further analyses on how kinematics of the movements relates to the prosodic rules.

The current results indicate that the prosodic rules in Polish can be flexible and adapt to the context of counting-out rhymes, even if it means breaking with the usual stress patterns. We suggest that integrating various speech tasks, registers, and contexts, including joyful rhythmic counting-out rhymes played in childhood, can broaden the perspective into potential gesture-speech interactions, revealing the stability and flexibility in speech prosody when gestures are involved. This evidence sheds light on the extent to which manual motoric gestures shape linguistic representations and emphasizes the need for more cross-linguistic work on speech-gesture interaction, particularly involving languages with diverse mora, tone, or stress.

Index Terms: gesture-speech coordination, Polish, counting-out rhymes, pointing, prominence

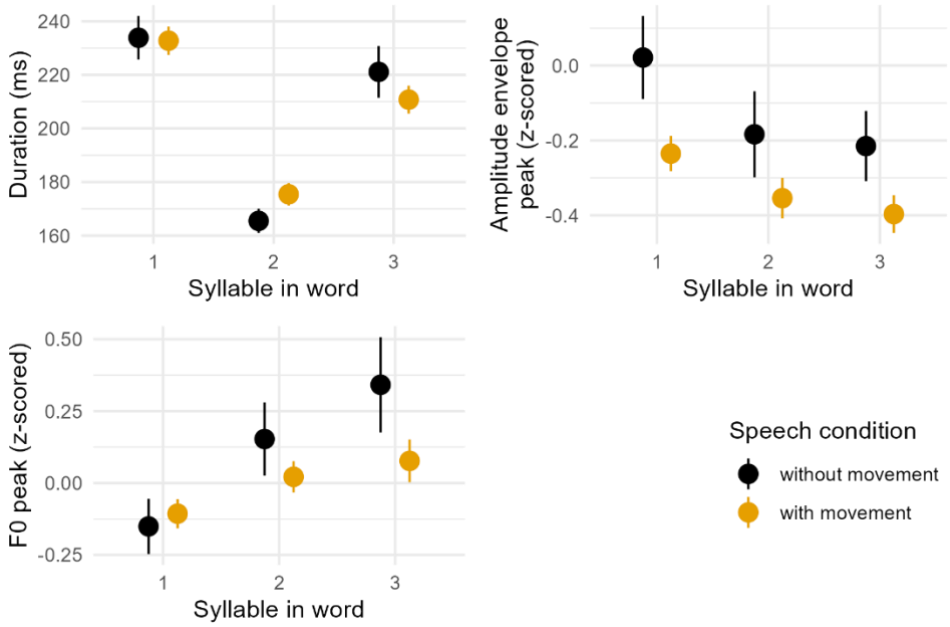

Figure 1: The figure illustrates one specific rhyme that only consists of three-syllabic words. As can be seen the three prosodic features are used differently, however in neither case the penultimate syllable is marked. In particular, the initial syllable receives most of the prominence marking, through duration and amplitude envelope.

References

Cutler, A. (2005). 11 Lexical Stress. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 264–289). Balckwell Publishing. https://repository.ubn.ru.nl/bitstream/handle/2066/56545/56545.pdf

Ćwiek, A., & Wagner, P. (2018). The Acoustic Realization of Prosodic Prominence in Polish: Word-level Stress and Phrase-level Accent. *9th International Conference on Speech Prosody 2018*, 922–926. https://doi.org/10.21437/SpeechProsody.2018-186

Dłuska, M. (1976). *Prozodia języka polskiego*. Państwowe Wydawnictwo Naukowe.

Dogil, G., & Williams, B. (1999). The phonetic manifestation of word stress. In *Word Prosodic Systems in the Languages of Europe*. Walter de Gruyter.

Franich, K. (2022). How we speak when we speak to a beat: The influence of temporal coupling on phonetic enhancement. *Laboratory Phonology*, *13*(1), Article 1. https://doi.org/10.16995/labphon.6452

Grimme, B., Fuchs, S., Perrier, P., & Schöner, G. (2011). Limb versus Speech Motor Control: A Conceptual Review. *Motor Control*, *15*(1), 5–33. https://doi.org/10.1123/mcj.15.1.5

Jassem, W. (1962). *Akcent języka polskiego* (Vol. 31). Wzdawnictwo Polskiej Akademii Nauk.

Kisler, T., Reichel, U., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, *45*, 326–347. https://doi.org/10.1016/j.csl.2017.01.005

Krivokapić, J., Tiede, M. K., & Tyrone, M. E. (2017). A Kinematic Study of Prosodic Structure in Articulatory and Manual Gestures: Results from a Novel Method of Data Collection. *Laboratory Phonology*, *8*(1). https://doi.org/10.5334/labphon.75

Malisz, Z., & Żygis, M. (2018). Lexical stress in Polish: Evidence from focus and phrase-position differentiated production data. *Speech Prosody 2018*, 1008–1012. https://doi.org/10.21437/SpeechProsody.2018-204

Osowicka-Kondratowicz, M. (2021). Wariantywność akcentowa w języku polskim. Proparoksytoneza. *Prace Językoznawcze*, *23*(4), 43–60. https://doi.org/10.31648/pj.7049

Osowicka-Kondratowicz, M. (2022). Wariantywność akcentowa w języku polskim. Oksytoneza. *Prace Językoznawcze*, *24*(2), 125–133. https://doi.org/10.31648/pj.7738

Peperkamp, S., Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics*, *38*(3), 422–430. https://doi.org/10.1016/j.wocn.2010.04.001

Rochet-Capellan, A., Laboissière, R., Galván, A., & Schwartz, J.-L. (2008). The Speech Focus Position Effect on Jaw–Finger Coordination in a Pointing Task. *Journal of Speech, Language, and Hearing Research*, *51*(6), 1507–1521. https://doi.org/10.1044/1092-4388(2008/07-0173)

Rubach, J., & Booij, G. E. (1985). A Grid Theory of Stress in Polish. *Lingua*, *66*(4), 281–320. https://doi.org/10.1016/0024-3841(85)90032-4

Stoltmann, K., & Fuchs, S. (2017). Syllable-pointing gesture coordination in Polish counting out rhymes: The effect of speech rate. *Journal of Multimodal Communication Studies. Special Issue: Gesture and Speech in Interaction*, *4*(1), 63–68.