Speech production is shaped by the lexical properties of words. Low-frequency words in dense phonological neighborhoods tend to be hyperarticulated, while high-frequency words in sparse neighborhoods tend to be hypoarticulated. Most research on this pattern has been drawn from English, with little evidence available for Polish. The question of whether modern AI speech systems reproduce these effects has likewise received little attention.

To address these gaps, this project examines lexically driven variation in vowel production in Polish across three domains: native human speech, synthesized speech, and speech recognition. Formant patterns were analyzed in the speech of native Polish speakers and in two commercial synthesis systems (Amazon Polly and ElevenLabs). The internal activations and confidence metrics of a speech recognition model (Meta's wav2vec 2.0) were then extracted to assess how these effects are encoded.

This research makes three contributions. The first extends the empirical base for lexical effects on phonetic production to Polish. The second evaluates whether commercial neural synthesis systems reproduce these effects. The third adapts existing methods to assess the degree to which speech recognition models reflect human perception. The findings have implications for cross-linguistic assumptions about lexical effects on speech production, the naturalness of speech synthesis, and the alignment of speech recognition systems with human perception.