Cechy leksykalne słów mają wpływ na artykulację mowy. Rzadziej występujące słowa oraz słowa należące do sąsiedztw fonologicznych o wysokiej gęstości zwykle podlegają hiperartykulacji, natomiast słowa częściej występujące i te będące w sąsiedztwach o niskiej gęstości – podlegają hipoartykulacji. Większość badań potwierdzających te tendencje dotyczy języka angielskiego, podczas gdy dane dotyczące języka polskiego pozostają nieliczne. Ponadto niewiele jest badań dotyczących tego, czy współczesne systemy sztucznej inteligencji potrafią odwzorować wspomniane zależności.

Niniejsza praca ma na celu zbadanie tych kwestii poprzez analizę zmienności artykulacji samogłosek, a konkretnie zmienności uwarunkowanej kontekstem leksykalnym w języku polskim, w trzech przypadkach: w przypadku mowy rodzimych użytkowników języka, mowy syntetycznej oraz systemu rozpoznawania mowy. W tym celu przeanalizowano strukturę formantów samogłoskowych w nagraniach rodzimych użytkowników języka polskiego oraz w nagraniach wygenerowanych za pomocą dwóch komercyjnych systemów syntezy mowy (Amazon Polly i ElevenLabs). Następnie wyodrębniono wewnętrzne aktywacje oraz metryki pewności modelu rozpoznawania mowy (wav2vec 2.0 firmy Meta), aby ocenić sposób kodowania wspomnianych zależności.

Badanie poszerza stan naszej wiedzy w trzech aspektach. Po pierwsze, badanie wzbogaca bazę empiryczną dotyczącą wpływu czynników leksykalnych na artykulację o język polski. Po drugie, badanie sprawdza, w jakim stopniu komercyjne systemy syntezy neuronowej potrafią odwzorować wspomniane wcześniej zależności. Po trzecie, w badaniu dokonano oceny stopnia, w jakim modele rozpoznawania mowy odzwierciedlają ludzką percepcję. Uzyskane wyniki mają znaczenie w kontekście uniwersalnych założeń językowych dotyczących czynników leksykalnych wpływających na artykulację mowy, a także w kontekście problemu stopnia naturalności mowy syntetycznej oraz zgodności systemów rozpoznawania mowy z percepcją ludzką.