



Streszczenie rozprawy doktorskiej

## **Opracowanie metodyki przetwarzania sygnałów akustycznych i danych obrazowych dla celów komputerowego wspomaganie diagnostyki logopedycznej z wykorzystaniem technik sztucznej inteligencji**

**Autor:** mgr inż. Agata Sage

**Promotor:** dr hab. inż. Paweł Badura, prof. PŚ

Niniejsza rozprawa podejmuje temat komputerowego wsparcia diagnostyki logopedycznej z wykorzystaniem metod sztucznej inteligencji. Zweryfikowano istnienie relacji pomiędzy normatywnymi i nienormatywnymi cechami głosek dentalizowanych a cechami wizualnymi obrazów i akustycznymi sygnałami prezentującymi mowę dzieci w wieku przedszkolnym.

Zaproponowano wieloetapową strukturę badań, która obejmowała: rejestrację bazy danych, opracowanie metodyki równoległego przetwarzania danych obrazowych i sygnałów akustycznych oraz przeprowadzenie testów statystycznych, które sprawdzały relacje wizualno-akustyczno-artykulacyjne. Badania bazowały na danych zarejestrowanych podczas sesji pomiarowych w przedszkolach. Materiał obejmował nagrania wideo i sygnał akustyczny mowy dzieci oraz opis logopedyczny przygotowany przez specjalistów terapii mowy.

Kolejny etap dotyczył przygotowania dwugałęziowej metody przetwarzania nagrań, osobno dla obrazów i dźwięku. W ramach ścieżki dotyczącej materiału wideo opracowano dwuetapową metodę segmentacji artykulatorów (warg, ust, zębów, języka): detekcji obiektów za pomocą sieci YOLO i ich segmentacji z wykorzystaniem modelu DeepLabv3+. Do wstępnego uczenia sieci zaproponowano wykorzystanie niedokładnych etykiet uzyskanych za pomocą metody zbioru poziomicy i obrazów rozmytych. W kolejnym kroku model został dostrojony za pomocą mniejszego zbioru obrazów z obrysami eksperckimi. Równolegle przetwarzano sygnał akustyczny tego samego fragmentu nagrania.

Na podstawie segmentacji wyekstrahowano parametry obrazowe: dwu- oraz trójwymiarowe. Zestaw parametrów obrazowych obejmował cechy teksturowe i parametry kształtu ust, warg i języka. W przypadku sygnałów akustycznych dla każdej ramki wyodrębniono zestaw cech, który obejmował cechy w dziedzinie czasu, cechy częstotliwościowe w pełnym pasmie i w pasmie szumu.

Wyniki analizy wykazały istnienie różnic między normatywnymi i patologicznymi realizacjami głosek w kontekście wybranych cech artykulacyjnych. Rezultaty sugerują największą użyteczność cech obrazowych bazujących na wolumenach, głównie opisujących kształt języka. Metodyka opisana w pracy oraz zaprezentowane rezultaty mogą stanowić punkt wyjścia do rozwoju systemów eksperckich wspierających diagnostykę logopedyczną sygnalizacją.

**Słowa kluczowe:** inżynieria biomedyczna, komputerowe wspomaganie diagnostyki logopedycznej, sygnalizacja, detekcja i segmentacja artykulatorów, uczenie głębokie