Doctoral thesis abstract

Analysis of spatial acoustic models of sibilant sounds in diagnosis of sigmatism

Author: mgr inż. Michał Kręcichwost

Supervisor: dr hab. inż. Paweł Badura, prof. PŚ **Auxiliary supervisor:** dr inż. Joanna Czajkowska

This doctoral dissertation addresses the use of computer speech signal processing to support speech therapy. The research concerns the analysis and classification of normative and pathological realization of sibilant consonants in preschool children. The research was based on the developed and collected speech database of 5 and 6-year-old children, containing acoustic data and a diagnostic description prepared by speech therapy specialists.

A measuring device for acquiring a 15-channel speech signal was designed, constructed, and tested. A multichannel acoustic signal processing methodology was developed based on deep learning. First, synchronization and signal preprocessing dedicated to sibilants were implemented. Data aggregation was performed using spatial signal processing techniques. Base feature extraction relied on the signal spectrum decomposition using a dedicated filter bank. A specific frequency range was proposed and compared to approaches involving the ranges reported in the literature. The advantage of linear-scale filter bank bandwidth extended towards higher frequencies was proven in the experiments. The data was collected into a three-dimensional structure of color images, defined as an acoustic volume. A three-branch convolutional neural network architecture was proposed for the data analysis. The goal of each processing branch was to search for patterns in separate dimensions: time, frequency, and time-frequency (mixed). Depending on the multichannel acoustic data interpretation, multiple models were developed and validated. Statistical analysis was performed for the 5-channel variant addressing the relationship between acoustic and articulatory features the normative and pathological realization of /s/ and /ʃ/ sibilants.

The conducted experiments over different variants of the proposed network structure indicate the deep learning-based models' ability to recognize various realizations of the analyzed sibilant sounds. Convolutional neural networks were additionally subjected to sensitivity analysis to investigate the impact of major parameters and settings on the models' performance. The experimental results legitimate the choices of proposed acoustic features, filter bank bandwidth and scale, measurement device configuration, signal processing procedures, as well as deep learning models in addressing the problem of detection of various realizations of sibilants under consideration.

Keywords: computer-aided speech therapy, sigmatism, speech signal, acoustic models, convolutional neural networks