THEMIS-SV: Automatic classification of language disorders from speech signals

Charalambos Themistokleous and Dimitrios Kokkinakis
Department of Swedish, University of Gothenburg, Sweden

Introduction
Brain injuries from stroke can affect the production of speech and can result in language impairments, such as aphasia (Brookshire and McNeil, 2014). Similarly, neurodegenerative conditions, such as Mild Cognitive Impairment and Alzheimer’s Disease, can affect speech and language. Here, we proposed the use of machine learning combined with automatic linguistic analysis to language samples produced by persons at various stages of cognitive decline in order to identify important speech markers. THEMIS-SV is a machine learning system that can be employed for the diagnosis of language disorders in Swedish that result from stroke or neural degeneration. A case study demonstrates the application.

Case Study

A. Transcription and Segmentation: The input of the system are transcribed recordings of speech. From these transcribed recordings of speech productions (see Figure 1).

B. Feature Extraction: We then extract acoustic features about the vowels and consonants, which serve as the basic input for machine learning models.

C. Model Selection/Evaluation: machine learning algorithms, e.g., support vector machines, random forests and neural network architectures are evaluated through model comparison. The outperforming models are those that are selected for the task of classification.

D. Classification: Final classification model.

Fig. 1: Phases of automatic speech analysis of Swedish (speech-to-text) using Machine Learning.

Procedure

A. Transcription and Segmentation: The input of the system are transcribed recordings of speech. From these transcribed recordings of speech productions (see Figure 1).

B. Feature Extraction: We then extract acoustic features about the vowels and consonants, which serve as the basic input for machine learning models.

C. Model Selection/Evaluation: machine learning algorithms, e.g., support vector machines, random forests and neural network architectures are evaluated through model comparison. The outperforming models are those that are selected for the task of classification.

D. Classification: Final classification model.

Fig. 2. Transcribed speech samples.

Case Study

The study aims to determine whether speech properties from vowels produced in a reading task by 55 Swedish speakers—30 healthy controls and 25 MCI—can distinguish MCI and healthy neurotypical speakers.

- 30 healthy controls and 25 MCI—between 55 and 79 years old (M=69, SD=6.4) participated in the study.
- The two groups did not differ with respect to age (t(52.72) = −1.8178, p = n.s.) and gender (W = 1567.5, p = n.s.), as is evident by the non-significant results from a t test and an independent 2-group Mann-Whitney U Test respectively.

Fig. 3. Training accuracy, validation accuracy, training loss, and validation loss for the Base Model, Model 1, Model 2, Model 3, and Model 4.

Results

D. Classification Results: Our results using acoustic measures: i.e., formant frequencies of vowels, fundamental frequency, and temporal measures provided high classification 85% classification accuracy on the unknown training data which is considerably higher than the classification accuracy achieved in earlier studies using speech features for the classification of MCI and healthy individuals.

Conclusions

- There has been considerable recent interest into machine learning applications for the study of speech produced by individuals with MCI and healthy controls (e.g., Kokkinakis et al. 2018, Fraser et al. 2018). Here we demonstrated an application, that enables the classification of MCI and healthy controls with high accuracy.
- The method proposed here can be employed for the analysis of speech of individuals with post-stroke aphasia and other language disorders and constitutes a promising step towards a fully automated differential diagnostic tool for language disorders.

References