Motif discovery in speech: application to monitoring Alzheimer's disease

Dragana Nikolić[1], Anna Barney[1], Vanda Nemes[2], Peter Garrard[2]

1. Institute of Sound and Vibration Research, University of Southampton, UK
2. Division of Clinical Sciences, St George's University of London, UK

Abstract.

INTRODUCTION: One of the most common presenting features of Alzheimer's disease (AD) is perseverative behavior: the tendency to make the same statement, ask the same question, or carry out the same action repeatedly over the course of the day. Since this phenomenon is widespread among patients with dementia, it is generally regarded as a sensitive indicator of a cognitive disorder that probably becomes more frequent with the progression of the condition. Having a means of measuring the occurrence of repeated speech episodes as the patients go about their daily lives could be used as a diagnostic aid and a tool for monitoring the condition. Continuous speech recording in a real-world setting would, however, violates privacy, can be prone to contamination by external noise and would be difficult to interpret without manually segmenting the various sources of recorded language (i.e. patient and interlocutors). Recording of energy fluctuations from the vocal apparatus through the bones of the skull (bone-conducted speech), however, enables us to collect data that derive exclusively from the patient. In this study a methodology to record and analyze bone-conducted speech using motif discovery techniques to identify, quantify and assess repeated speech segments from large sets of recorded data, was proposed. In order to evaluate the performance of the adopted method for speech pattern detection, pilot data consisting of both air- and bone-conducted speech recorded by actors using scripted texts with certain short questions and statements embedded several times, have been collected and processed and the preliminary results obtained are reported.

METHOD: A proof of concept experimental testing was undertaken on 5 healthy subjects. The subjects were instructed to read aloud 3 predefined scripts containing short, embedded, repeated questions and statements. Bone-conducted speech was recorded using an accelerometer attached to the skin above the temporo-mandibular joint, amplified and sampled at a frequency of 16 kHz. Synchronous recording of the air-conducted speech via a conventional headset microphone served as a reference to validate the accelerometer data. The accelerometer data, devoid of any semantic content, was automatically analyzed using signal processing techniques. Speech segments were extracted from the recordings, bandpass filtered, denoised and divided into frames for feature extraction. A set of 95 features, including statistical measures, spectral moments, Mel-frequency cepstral coefficients, Perceptual Linear Predictive Cepstral Coefficients and prosodic measures, were calculated for each frame. Principal component analysis was then applied to reduce dimensionality of the

feature vector and only the first principal component was retained for further processing. Motif detection was undertaken using Symbolic Aggregate Approximation conversion to transform the data in a way that made manipulation more convenient using symbolic words and organizing them into buckets containing all subsequences of time-series with the same hash values. Thus, searching for a motif was performed within those smaller groups of subsequences rather than on the whole time series. This rearrangement of the data is beneficial as it narrows and accordingly speeds up the search. A search performed through smaller sections of a similarity matrix based on the Approximate Distance Map algorithm was then used to identify candidate motifs likely to be associated with repeated patterns. An alignment-based clustering of the detected motifs was carried out at the end in order to group similar but not entirely aligned motifs.

RESULTS: The approximately 4 minute long recordings with three embedded motifs repeated 4, 9 and 14 times were processed using the method described. Temporal locations of the detected patterns were then compared with those of the true motifs marked by listening to the simultaneously recorded air-conducted speech. The correct rate of the detection was between 50.00-71.43% and the precision was between 62.50-90.91% depending on the subject.

CONCLUSION: The initial results indicate that the adopted approach has potential for detecting repetitive speech patterns. Further improvements are underway including fine-tuning of the algorithm parameters with the aim of improving the detection rate and precision.

In the next stage of the study, we will analyze bone-conducted speech recorded in patients with possible or probable AD to establish the feasibility and tolerability of the proposed methodology in a natural environment. Future work will comprise larger subject populations with the aims of: defining distinct patterns of perseverative speech; evaluating perseveration rates and correlating them with stage and progression in dementia; and calibrating treatment effects from established interventions. If successful, the technique could form the basis of a device for monitoring performance in, for instance, trials of disease modifying drugs for AD.