Mobile Application for Real Time Monitoring of Speech Rate Based on Phoneme Segmentation Techniques

Katia Raichlin Levi [1], Aviv Sotzianu [2], Ofer Amir [3], Eran Aharonson [2], Zehava Ovadia-Blechman [1]

[1] Medical Engineering Department, Afeka Tel-Aviv Academic College of Engineering.
[2] Software Engineering Department, Afeka Tel-Aviv Academic College of Engineering.
[3] Communication Disorders Department, Tel Aviv University.

ekatirinar@mail.afeka.ac.il, avivs4@afeka.ac.il,oferamir@post.tau.ac.il, erana@afeka.ac.il, zehava@afeka.ac.il

Abstract Stuttering is a prevalent speech disorder. It is characterized by various speech disfluencies, such as word repetitions, syllable repetitions, prolongation of sounds and blocking or hesitation before word completion**.** Although, among adults, stuttering is not regarded as a disorder that can be cured by therapy; adults-who-stutter can learn various techniques to control and improve their speech fluency. To that end, they are commonly required to either slow or modify their speaking rate, and become aware of it, as they produce speech spontaneously. A known method to quantify client's speaking rate is by calculating the number of syllables or phonemes during practice within a fixed time.

The goal of the present study was to develop an algorithm, as well as an application for smartphone, that monitors in real time the rate of the speech of a patient and presents him the results both graphically and numerically in real time. The recordings and their analysis are stored on a cloud. This data is uploaded to therapist's website which allows the therapist to send his patients real-time feedback and instructions. Moreover, in addition to the patient, the therapist can also track training results through a dedicated website.

The solution is based on a digital speech processing algorithm which receives speech signal, extracts relevant features and calculates the rate of the user's speech by recognizing the number of phonemes in each time segment. The analyzed data is presented graphically and numerically on the screen. The DSP algorithm was written using Matlab and ported to Java implementation on Android OS based smartphone.

Our algorithm is language agnostic, for this research we analyzed spontaneous Hebrew language speech in real time.

The developed algorithm divides user's speech signal to small overlapping speech frames. The algorithm applies Hamming window for each frame and extracts Mel-Frequency Cepstrum Coefficients (MFCC) in each one. Subsequently, Spectral transition measure (STM) is calculated for each frame. STM step outputs possible candidates that represent the possible transition between two adjacent phonemes. The algorithm detects and removes false boundaries and calculates out of it the average number of phonemes spoken per minute.

The preliminary results show that accuracy count phonemes reach an average of 90% (compared to human phoneme counting). The preliminary analysis shows that the main source of deviation is background noise.

In Summary, our novel application was found reliable and has a clinical potential to improve speech therapy. The algorithm providing an efficient tool for patient who enables real-time feedback during his exercises, and allowing a better effective follow of the speech fluency.

1. Introduction

Stuttering is a prevalent speech disorder. It is characterized by various speech disfluencies, such as word repetitions, syllable repetitions, prolongation of sounds and blocking or hesitation before word completion**.**
Although, among adults, stuttering is not regarded as a disorder that can be cured by therapy; adults-who-stutter can learn various techniques to control and improve their speech

fluency. To that end, they are commonly required to either slow or modify their speaking rate, and become aware of it, as they produce speech spontaneously. Today acceptable clinical practice to quantify the client's speech rate is done by manually calculating the number of phonemes during training within a fixed time.

Speech therapists can use the following stuttering therapy techniques:
Auditory feedback devices:
Auditory feedback devices measure feedback speech, distorted speech, delayed speech or pulse tones. An example to such devices is the Electronic fluency devices which change the sound of the user's voice in his or her ear. The disadvantage of those devices is that they are available to the client only during the therapy session.
Computer based therapy:
Computer based therapy programs display speech spectrograms, waveforms, pitch patterns and other graphical representations of an individual's speech. They are available only in large clinical facilities. The training process can take many months of repeated procedures that are costly.

Using those techniques, the patient can evaluate his speech rate only during speech therapy sessions. However, between the meetings he can't be sure whether his exercises improve his speech rate.

The goal of the present study was to develop an application for smartphone which monitors, in real time, the rate of the speech of a patient and presents the results both graphically and numerically. Moreover, in addition to the patient, the therapist could also enable to track training results through a dedicated cloud service.

2. Methodology

2.1 Speech Rate Algorithms and Applications

The following algorithms/applications evaluate speech rate to the following groups:

Algorithms for detection of phonemes boundaries for evaluation of speech rate:
Example for such algorithm is Dusan and Rabiner algorithm [11].This algorithm wasn't implemented on mobile, it doesn't work in real time and it wasn't tested on Hebrew speakers.

Algorithms which measures speech rate by counting vowels in time unit:
Example for such algorithm is Pfau and Ruske [**Error! Reference source not found.**]. This algorithm doesn't work in real time and it wasn't implemented on mobile.
Evaluation of speech rate by counting the number of syllables in time segment:
Example for such software is AgentAssist (Real Time Speaking Rate Monitoring System). This software was designed for call centers which are interested to learn whether their representatives speak too fast [3].It analyses voice signal in real time, however it wasn't implemented yet on mobile.

2.2 DSP algorithm

The DSP algorithm (Figure 1) is based on Sorin Dusan & Lawrence Rabiner algorithm [1] for

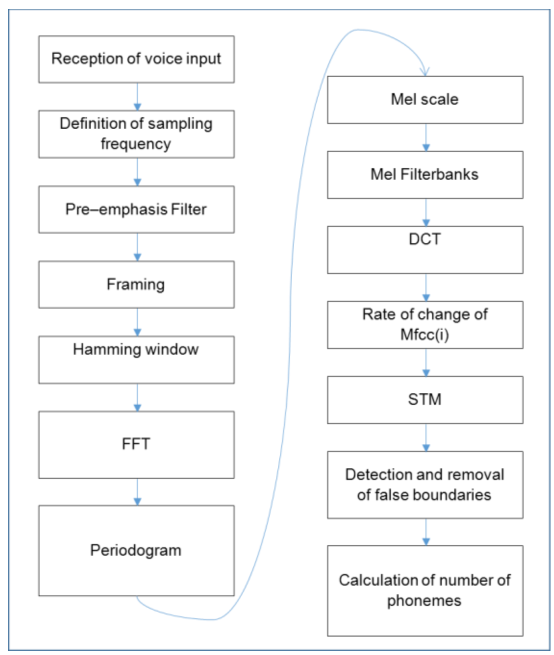

Figure 1: Block Diagram of the DSP algorithm

detection of phonemes boundaries. However, we have modified the detection and removal of false boundaries stage of the algorithm. The DSP algorithm was written using Matlab. The sample rate of the algorithm is 16 KHz. The algorithm divides user's speech signal to time segments of 10 seconds with 1 second overlap (chunks). It uses the process chunk function in order to calculate the number of phonemes in every chunk.

The voice signal in each chunk is divided to small speech frames (32 ms) with overlap (22 ms). The algorithm applies Hamming window for each frame and extracts 10 Mel-Frequency Cepstrum Coefficients (MFCC) in each one. Subsequently, Spectral transition measure (STM) is calculated for each frame. STM outputs possible candidates that represent the possible transition between two adjacent phonemes.

In order to detect the false boundaries, we defined a threshold for the STM: if the current STM is smaller than 0.25, it is recognized by the algorithm as a false boundary and is removed. If the STM is higher than 0.25, it is considered as a phoneme boundary.

We calculate the number of phonemes in each chunk by the following equation: Number of phonemes = number of boundaries – 1.

We consider the last chunk as a special case: If the number of samples in the last chunk<16000 samples, we ignore the last chunk. If the number of samples in the last chunk is higher than 16000 samples, we consider the last chunk and calculate the number of phonemes in it.

2.3 The Protocol

In order to test the accuracy of the algorithm, we have recorded 18 random subjects (men and women), Hebrew speakers, without speech disorders, age 20-50. The subjects were asked to read a Hebrew text from the book:"Habait shel Yael". We compared the number of phonemes, which were detected by the application, to manual count of phonemes.

2.4 Smartphone and Tablet applications

We have developed a complete system consisting from a smartphone application (Android) for the patient and tablet application (Android) for the speech therapist.

2.4.1 Smartphone (Android) Application

The smartphone (Android) application monitors in real time the rate of the speech of a patient and presents him the results both graphically and numerically (Figure 2). The recordings and their analysis are stored on a cloud. The data is uploaded to therapist's website (or tablet) which allows the therapist to send his patients real-time feedback and instructions. The application updates the graph of speech rate every 2 seconds, based on arithmetic mean.

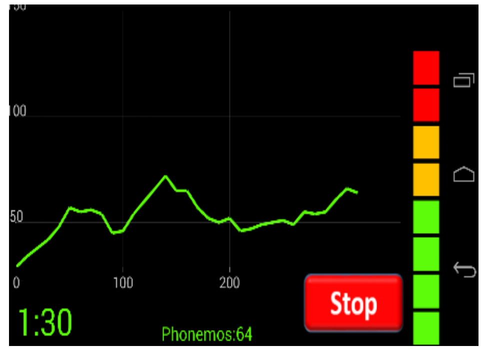

Figure 2: Online analysis on smartphone

3. Preliminary Results

Figure 3 presents the accuracy of the algorithm:

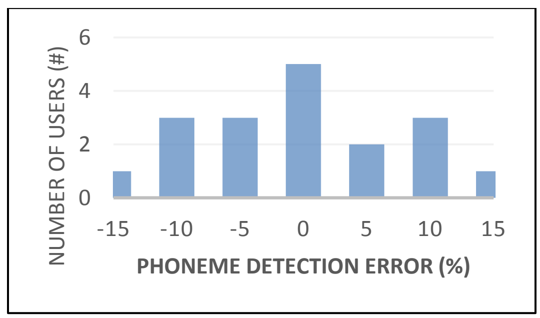

Figure 3: Algorithm Accuracy Histogram

If the algorithm detects fewer boundaries than the manual count, the accuracy is negative. If the algorithm detects more boundaries than the manual count, the accuracy is positive.

The preliminary results show that accuracy of the algorithm is 92.74% (compared to manual count). The following parameters influenced algorithm's accuracy: environmental noise, the proximity of the microphone to the speaker, too fast speech rate, the gender of the speaker etc.

4. Discussion

Currently, patients can evaluate their speech rate only during speech therapy sessions. Therefore, a need arises for a mobile application which would monitor in real time speaker's rate of speech in order to improve his speech fluency. On the market, there isn't yet any mobile application which monitors speech rate in real time based on counting the number of phonemes in time interval, like our application. With further tests and improvements, this application has the potential to improve speech therapy process.

5. Summary

This article describes smartphone (Android) application which monitors the speech rate of a patient and presents, in real time, the results both graphically and numerically. Our algorithm and application has the potential to improve the speech therapy by providing the patient real-time feedback and instructions on his exercises by his speech therapist, thus allowing him to learn to control his speech fluency. This application may be effective in different cases where the need to improve the rate of speech. Such as slow or increase the rate of speech, control the rate of uniform speech. Furthermore, the application may also be useful for practicing the speech volume control.

Recommendations for further work

Our application has been tested on patients without speech disorders. In order to confirm that our application suits their needs, it is highly recommended to test the application with those patients.

It is recommended to test the accuracy of the application in different rates of speech: slow, normal and fast.

The application is currently designed for adults. With some modifications it could be used also in children's therapy.

References

1. Sorin Dusan, and Lawrence R. Rabiner. "*On the relation between maximum spectral transition positions and phone boundaries*." *INTERSPEECH*. 2006.
2. Thilo Pfau, and Günther Ruske. "*Estimating the speaking rate by vowel detection*." *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*. Vol. 2. IEEE, 1998.
3. Pandharipande, Meghna Abhishek, and Sunil Kumar Kopparapu. "*Real time speaking rate monitoring system.*" *Signal Processing, Communications and Computing (ICSPCC), 2011 IEEE International Conference on*. IEEE, 2011.
4. Gerald A. Maguire, Christopher Y. Yeh, and Brandon S. Ito, "*Overview of the diagnosis and treatment of stuttering*" *Journal of Experimental & Clinical Medicine* 4.2 (2012): 92-97. Urmila Shrawankar and V. M. Thakare. "*Techniques for Feature Extraction in Speech Recognition System: A Comparative Study*", International Journal Of Computer Applications In Engineering, Technology and Sciences (IJCAETS),ISSN 0974-3596,2010,pp 412-418
5. Kalinli, Ozlem. "*Automatic Phoneme Segmentation Using Auditory Attention Features*." *INTERSPEECH*. 2012.
6. Amir Ofer, and Reut Levine-Yundof. "*Listeners' attitude toward people with dysphonia*." *Journal of Voice* 27.4 (2013): 524-e1.