Journal of Fluency Disorders 51 (2017) 60–68

Contents lists available at ScienceDirect

Journal of Fluency Disorders

Short communication

A real-time phoneme counting algorithm and application for speech rate monitoring

Vered Aharonson[a,b,*], Eran Aharonson[c], Katia Raichlin-Levi[a], Aviv Sotzianu[c], Ofer Amir[d], Zehava Ovadia-Blechman[a]

[a] Medical Engineering Department, Afeka Tel Aviv Academic College of Engineering, Israel
[b] School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, South Africa
[c] Software Engineering Department, Afeka Tel Aviv Academic College of Engineering, Israel
[d] Communication Disorders Department, Tel Aviv University, Israel

ARTICLE INFO

Article history:
Received 15 June 2016
Received in revised form
15 December 2016
Accepted 13 January 2017
Available online 15 January 2017

Keywords:
Stuttering therapy
Speaking rate computation
Spectral transition measure
Smartphone application

ABSTRACT

Adults who stutter can learn to control and improve their speech fluency by modifying their speaking rate. Existing speech therapy technologies can assist this practice by monitoring speaking rate and providing feedback to the patient, but cannot provide an accurate, quantitative measurement of speaking rate. Moreover, most technologies are too complex and costly to be used for home practice. We developed an algorithm and a smartphone application that monitor a patient's speaking rate in real time and provide user-friendly feedback to both patient and therapist. Our speaking rate computation is performed by a phoneme counting algorithm which implements spectral transition measure extraction to estimate phoneme boundaries. The algorithm is implemented in real time in a mobile application that presents its results in a user-friendly interface. The application incorporates two modes: one provides the patient with visual feedback of his/her speech rate for self-practice and another provides the speech therapist with recordings, speech rate analysis and tools to manage the patient's practice. The algorithm's phoneme counting accuracy was validated on ten healthy subjects who read a paragraph at slow, normal and fast paces, and was compared to manual counting of speech experts. Test-retest and intra-counter reliability were assessed. Preliminary results indicate differences of −4% to 11% between automatic and human phoneme counting. Differences were largest for slow speech. The application can thus provide reliable, user-friendly, real-time feedback for speaking rate control practice.

Introduction

Stuttering is found in 1% of the adult population speakers. It is characterized by various speech disfluencies, such as word repetitions, syllable repetitions, prolongation of sounds and blocking or hesitation before word completion (Duchin & Mysak, 1987; Maguire, Yeh, & Ito, 2012; Wingate, 1976). Although stuttering is not regarded as a disorder that can be cured by therapy (Lutz & Mallard, 1986), adults who stutter can learn various techniques to control and improve their speech fluency

* Corresponding author at: School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, South Africa.
E-mail addresses: vered@afeka.ac.il (V. Aharonson), vered.aharonson@wits.ac.za (E. Aharonson), ekatirinar@mail.afeka.ac.il (K. Raichlin-Levi), avivs4@afeka.ac.il (A. Sotzianu), oferamir@post.tau.ac.il (O. Amir), zehava@afeka.ac.il (Z. Ovadia-Blechman).

http://dx.doi.org/10.1016/j.jfludis.2017.01.001

(Kalinowski, Armson, Stuart, & Gracco, 1993; Starkweather, 1987). 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 (Andrade, Cervone, & Sassi, 2003; Duchin & Mysak, 1987; Kalinowski, Stuart, Sark, & Armson, 1996). Routine clinical practice to quantify the patient's speech rate during practice is currently done by manually calculating the number of syllables or phonemes within a fixed period of speech.

Several stuttering therapy technologies and device have been proposed and implemented. These devices aim to assist the patient in controlling stuttering by providing him/her with feedback. One type of device uses altered auditory feedback methods that delay or change the sound of the user's voice or play pulse tones and provide these sounds as feedback to the stutterer (Hargrave, Kalinowski, Stuart, Armson, & Jones, 1994; Kalinowski et al., 1996; Stuart, Kalinowski, Rastatter, Saltuklaroglu, & Dayalu, 2004). Another type of device uses a visual feedback method, in which computer programs display speech spectrograms, waveforms, pitch patterns and other graphical representations of an individual's speech (Awad, 1997; Hudock et al., 2011; Ingham et al., 2001). The disadvantage of most feedback devices is that they are designed for professional usage. Due to their high cost and complexity, these devices are mainly used in clinical facilities. They are therefore available to the patient only during clinical therapy sessions. Moreover, none of the existing devices provide the patient with *quantitative* feedback of speaking rate information.

Speaking rate estimation has various applications in speech recognition and speech synthesis technologies and its calculation is performed using different speech units (Morgan & Fosler-Lussier, 1998; Morgan, Fosler-Lussier, & Mirghafori, 1997; Ramus, 2002; Shrawankar & Thakare, 2013; Siegler & Stern, 1995; Verhasselt & Martens, 1996). The most common units are syllables, vowels or phonemes (De Jong & Wempe, 2009; Pellegrino, Farinas, & Rouas, 2004; Pfau & Ruske, 1998; Wang & Narayanan, 2007; Xie & Niyogi, 2006), or their combinations (Pfitzinger & Itzinger, 1998). Various digital signal processing (DSP) algorithms have been proposed to detect these speech unit boundaries (Dusan & Rabiner, 2006; Grayden & Scordilis, 1994; Ziolko, Manandhar, & Wilson, 2006).

To be effective in speech therapy feedback applications, these algorithms should, however, be implemented in real time. Although some real-time speech rate estimation has been implemented for different applications, like monitoring speaking rate for call center agents (Pandharipande & Kopparapu, 2011), automated voice response systems (Obaidat, Sevillano, & Filipe, 2012) and speech modification (Kupryjanow & Czyzewski, 2010), none were designed or employed for speech disorder analysis or therapies that require enhanced accuracy. Moreover, none of these methods have been implemented in applications for home use or for speaking rate control practice.

The goal of the present study was to develop a user-friendly mobile application that could provide accurate speaking rate feedback, and which could be used by patients for home practice, between their speech therapy sessions at the clinic. The device should provide a practice program tool for the patient, as well as program monitoring and a management tool for the therapist. The application proposed in this study is based on phoneme counting, is implemented on a smartphone and/or tablet and provides real-time feedback to the user.

Methodology

Design overview

Our system design is focused on usability for patients and their therapists, for home practice as well as for practice program supervision. The design entails a user-friendly interface and real-time feedback for the patient, as well as reliable and accurate speaking rate computation. Speaking rate is computed using a DSP algorithm that counts phonemes in pre-defined speech segments. The algorithm is implemented in real time on an Android application for mobile devices (smartphones or tablets). The input of the computation process is speech recorded by the mobile device's microphone and its output is continuously displayed on the mobile device's screen in a user-friendly graphical interface.

2.2. Speaking Rate Computing Algorithm

The algorithm is based on Dusan and Rabiner's algorithm (Dusan & Rabiner, 2006): The speech signal is pre-emphasized using a second-order, high-pass, infinite impulse response (IIR) filter, in order to emphasize rapid changes in the speech signal (Oppenheim, Schafer, & Buck, 1989). The signal is then segmented into 32 ms frames, with a 10 ms overlap, using a Hamming window. The spectrum of each frame is calculated using a periodogram. The spectrum's frequency axis is subjected to a log-based transform (mel-frequency scale), implemented by filter banks, and is then decorrelated using a modified discrete cosine transform (DCT). Ten Mel-frequency cepstrum coefficients (MFCC) as well as their derivatives (rate of change) are extracted (Huang, Acero, Hon, & Foreword By-Reddy, 2001), and spectral transition measure (STM) is calculated for each frame. The STM provides candidates for transition between two adjacent phonemes. A threshold algorithm determines for each candidate whether it is a real transition and if so, defines it as an inter-phoneme boundary. Speaking rate is calculated as the number of phonemes (or the number of inter-phoneme boundaries) per minute. A block diagram of the algorithm implementation is illustrated in Fig. 1.

The speech recording is processed in units of 10 s., and a phoneme count is obtained for each unit. This processing unit length is suitable for real-time implementation where the smartphone uses 10-s. buffers. A one-second overlap between