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
0094-730X/

(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

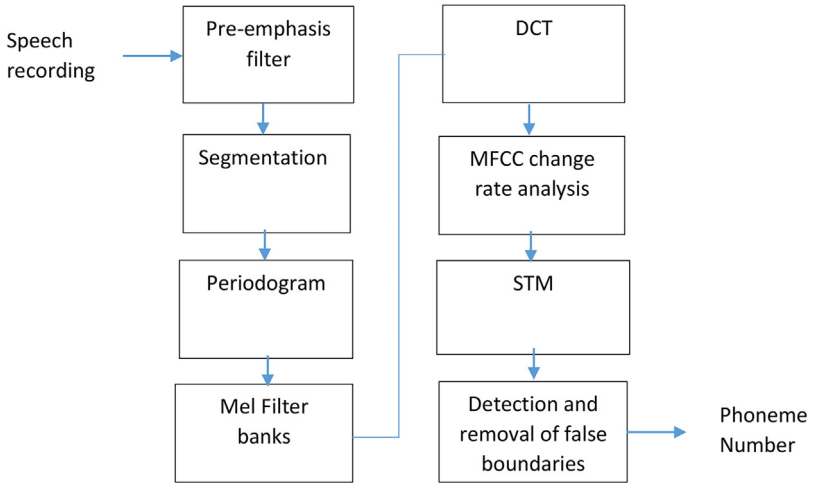

Fig. 1. Block diagram of the phoneme counting algorithm. The speech signal is segmented into frames. Pre-emphasis of each frame is followed by a spectrum calculation using a periodogram. The spectrum's frequency axis is subjected to a log-based transform (mel-frequency scale), implemented by filter banks, and is then decorrelated using a modified discrete cosine transform (DCT). Mel frequency cepstrum coefficients (MFCC) as well as their derivative (rate of change) are extracted and spectral transition measures (STM) are computed to produce candidate inter-phoneme boundaries. False boundaries are rejected using an adaptive threshold algorithm.

consecutive 10-s units is applied in order to smooth out the truncations. The last unit is omitted if its number of samples is smaller than 5 s. (8×10^5 samples), making it too short for this processing.

Our algorithm enhances the original STM algorithm in the detection and removal of false boundaries between adjacent phonemes. False boundaries are defined by an adaptive, signal-to-noise ratio (SNR)-based threshold for the STM values: Only values above this threshold are considered real transitions. The algorithm was developed in Matlab® and was ported to Java real-time implementation on Android OS, for both smartphone and tablet.

Smartphone application

The system consists of a smartphone application for the patient and a tablet application for the speech therapist. Both applications were programmed using Eclipse® (Android SDK plugin), using Photoshop® to facilitate the graphical user interface (GUI) design. The web server application was developed in the PHP language (Welling & Thomson, 2004). The choice of all implementation platforms was based on their wide-spread usage and their low cost, to enhance the application's usability.

The DSP algorithm was implemented in real time and was imported to the smartphone to compute the number of phonemes in 10-s buffers. The application updates the results every 2 s. The results are presented both graphically and numerically. The display also presents the speech volume. The screen is refreshed every 2 s to display a new number of phonemes and a new position of the graph. The layout of the GUI is displayed in Fig. 2.

All recordings and their analysis are stored on a cloud. The data are uploaded to the therapist's website and are handled by a tablet application which allows the therapist to send his patients real-time or off-line feedback and instructions.

System Performance Evaluation

The system performance evaluation tested both the phoneme counting accuracy of the algorithm and the real-time performance of the smartphone application. The evaluation experiment is described in the following sections.

Speech Recordings

The recordings for the current research consisted of read Hebrew speech. The text chosen was a paragraph from the "Thousand Islands" reading passage, a phonemically balanced Hebrew text, which has been used in previous speech therapy studies (Amir & Levine-Yundof, 2013). The written text contains six short sentences and a total of 280 phonemes. The text was read by ten speakers, five males and five females. All subjects were native Hebrew speakers between 22 and 30 years old, and none had speech disorders. All subjects signed an informed consent and the experiment protocol was approved by the Afeka Tel Aviv Academic College of Engineering Ethics Committee (ID 08-12-2015-2-AFK).

The recordings were performed in a quiet room. The subjects sat down and were given a smartphone with a minimal recording application, which displayed only a "Record" button that changed into a "Stop" button when pressed, and a printed page containing the "Thousand Islands" passage. The experimenters provided identical instructions to all subjects.

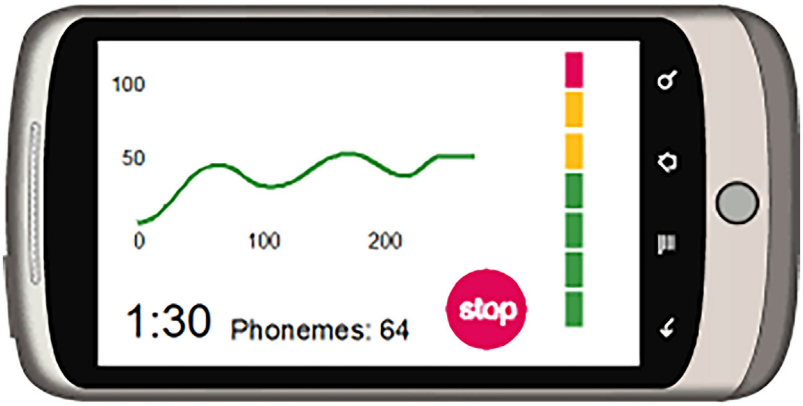

Fig. 2. Smartphone graphical and numeric display of phoneme count and speech intensity. A color graph of red, yellow and green on the right-hand side of the screen displays intuitive speech intensity ranges, where the user is required to keep the intensity in the green range. A green line graph on the left-hand side of the screen depicts the real-time phoneme count per minute. The instantaneous phoneme count is updated below this graph. A recording time display and a "stop" button are provided at the bottom of the screen. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The instructions (translated from Hebrew) were as follows: "Please read the text silently. Then take the smartphone, hold it at a comfortable distance of 30 cm, press the "Record" button and read the text out loud at your convenient, natural pace. Then read it aloud again but as **slowly** as you can. Lastly, read it aloud again but as **fast** as you can. When you finish reading, press the "Stop" button". A second version of the instructions, swapped the order of the latter two sentences and instructed the subject to read in a fast reading pace before the slow one. Two female subjects and three male subjects were randomly chosen by the experimenter to receive this second instructions version. The recordings were stored in an uncompressed "wav" file format.

Performance Evaluation

The consistency between the algorithm's Matlab® implementation and its smartphone real-time implementation was initially assessed by comparing phoneme counts in all 10-s units of speech.

The accuracy of the algorithm was tested by comparing its phoneme counting to that of two expert human listeners: a professional linguist and a professional speech researcher. Each listener counted the phonemes in the full text recordings as well as in all the 10-s units. The order of slow, normal and fast speech files in the counting sessions was randomly chosen. A week after the first counting session, each listener performed a second counting session. In this session a subset of ten recordings was randomly chosen, containing all reading rates, as well as a full recording session of one subject. One listener's counting results were then compared to the automatic counting of the algorithm. This listener was chosen by his better test-retest performance. All differences in counting were transformed to percentage units in order to simplify the comparison between different counting values.

Choosing a proper statistical test for test-retest reliability of each listener and for the agreement between the human listeners' counting and that of the algorithm is a subtle task (Bland & Altman, 1986): inter-rater agreement indices, such as Cohen's Kappa (Cohen, 1968), are suitable only for categorical, nominal rating scales, whereas indices for correlations between interval data measurements, like Pearson's product-moment "R", can indicate the association between two variables, but do not provide information about agreement (Müller & Büttner, 1994). We chose to present the agreement using Interclass correlation based on two-way ANOVA, denoted ICC(2,1) (Shrout & Fleiss, 1979). This analysis was performed for Test-Retest, Listener1-Listener2 and Listener-Algorithm comparisons, and for the latter, was complemented by a graphical Box-and-whiskers representation. The quality of the agreement manifested by the ICC values was evaluated using Cicchetti's criteria (Cicchetti, 1994; Fleiss, Levin, & Paik, 2013).

Differences between mean phoneme counts across speaking rates (i.e., normal vs. fast and normal vs. slow) were evaluated with separate *t*-tests. Statistical significance for all statistical analyses was defined as $p < 0.05$.

Results

The STM analysis of the algorithm is illustrated in Fig. 3. An input unit of 10 s of speech is presented in the top graph, followed by its STM computation in the bottom graph, where the circles represent transitions.

Comparison of the phoneme count between the two implementations of the algorithms – the offline count in Matlab® and the real-time count in the Android application – yielded identical counting for both implementations, for all speech files. Measurements of the response time of the Android implementation of the algorithm yielded a maximum delay of 3 s.

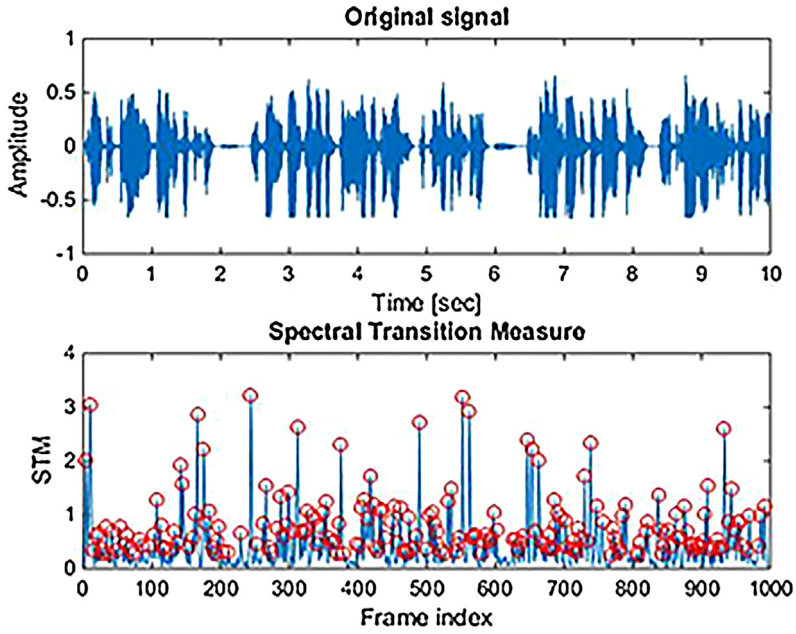

Fig. 3. An example of spectral transition measure (STM) computation: an input signal chunk (10 s) in the top graph and its STM per frame in the bottom graph.

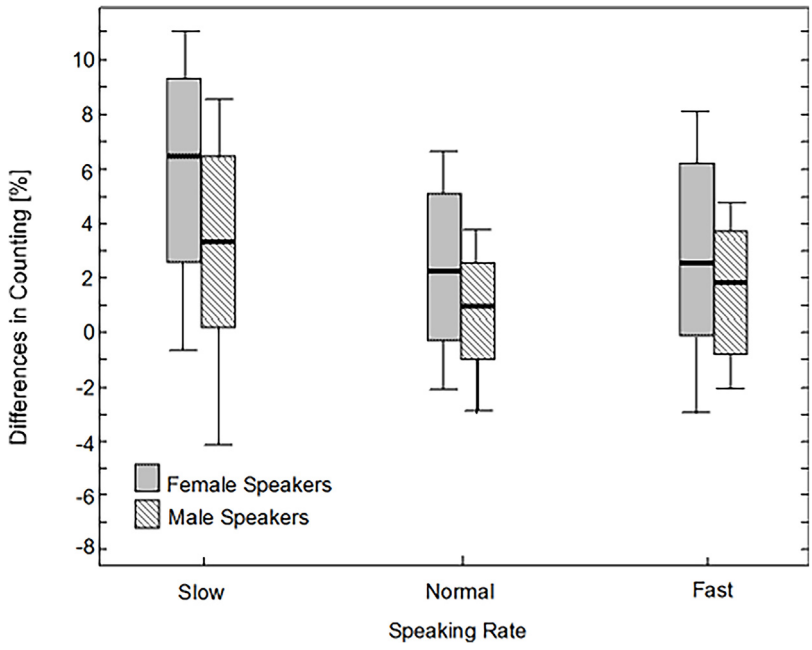

Fig. 4. Box-and-whisker plot of the differences between phoneme counting by the algorithm and Listener 2 for the three speaking rates. The difference values are normalized and presented as a percentage. Difference values for recordings by male and female speakers are denoted by dots and line textured boxes, respectively.

The majority of differences in counting between the first and the second sessions for both listeners were small; ranging between −1% and 1.5%, with rare outliers of up to 3%. The Intraclass correlation ICC(2,1) coefficients between test and retest counting were 0.992 for Listener 1 and 0.998 for Listener 2. The ICC(2,1) coefficient between the counting results of Listener 1 and Listener 2 was 0.991. These ICCs represent excellent agreement beyond chance. The phoneme counting of Listener 2, who had better test-retest performance, was chosen as the "gold standard" for comparison to the algorithm's phoneme counting accuracy. The differences between the counting of the algorithm and Listener 2 can thus be regarded as errors.

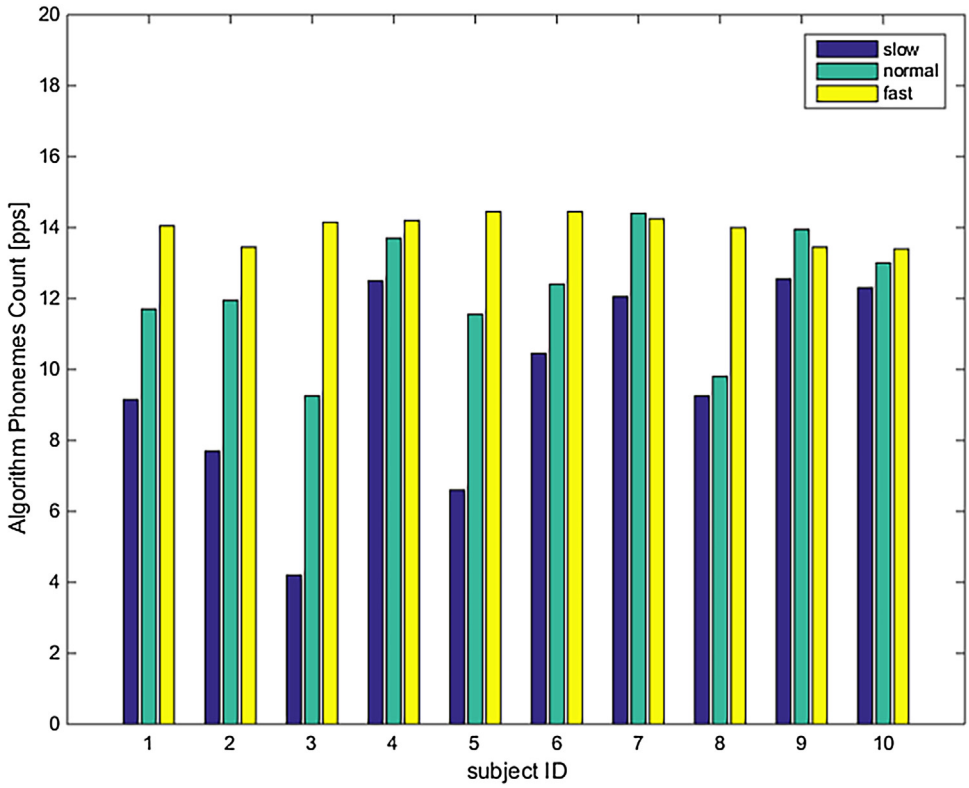

Fig. 5. Mean phoneme counting, in phonemes per second (PPS) of the algorithm for slow (blue), normal (green) and fast (yellow) rates for the ten subjects. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 4 illustrates the algorithm's errors distribution. The distribution's range, 50th percentile and median are presented in a box-and-whisker plot for each speaking rate and for male speakers' and female speakers' recordings. The figure shows that the majority of differences in counting results between the algorithm and the human listener are positive, indicating over-counting of the algorithm. Larger errors are indicated in the algorithm's phoneme counting of female speakers' recordings and for the slow speaking rate recordings. The distribution of errors for the fast speaking rate is similar to those of the normal speaking rate. These two distributions are more symmetrical between over- and under-counting of the algorithm (positive and negative values, respectively) for recordings of male speakers. The ICC(2,1) coefficient between the counting of Listener 2 and the automatic counting was 0.912, and represent good agreement beyond chance.

Fig. 5 portrays the mean phoneme counting of each subject, for the slow, normal and fast speaking rates, calculated across all 10-s units. The speaking rates are presented in phonemes per second (pps). While the mean of the fast speaking rate is rather similar for all subjects (around 14 phonemes pps, the normal and slow rates vary considerably among subjects. The extent of the decrease in mean phoneme count between the normal and slow speaking rates and of the increase in mean phoneme count between the normal and fast speaking rates varies considerably among the subjects, from as little as 6% to as much as 51%. Except for two subjects, whose mean phoneme count in normal speech was slightly higher than in fast speech, the mean phoneme counts were lower for the slow rate and higher for the fast rate compared to the normal rate. The differences between the pairs of speaking rate series (normal-slow and normal-fast) were statistically significant: $t(58) = 2.96$, $p = 0.021$ for normal vs. slow speaking rate, and $t(43) = -1.93$, $p = 0.037$ for normal vs. fast speaking rate.

Discussion

The system developed in this study is a low-cost, user-friendly practice tool that can aid patients with speech fluency disorders to control and/or alter their speaking rate. This application can enable the patients to practice at home and aid the clinician to monitor their patients' progress and enhance their therapy planning.

The system incorporates an algorithm that performs phoneme counting and a mobile (smartphone) application that implements this algorithm to compute, in real-time, a speaker's rate of speech. The system is novel both in its computation method, counting phonemes based on STM analysis, and its implementation in a real-time, user-friendly smartphone application that enables comfortable usage and uncomplicated interpretation of the results for the patient as well as providing an informative monitoring tool for the speech therapist.

The study incorporated a feasibility test for automated phoneme counting using the STM algorithm, with an adaptive threshold. The test included comparison of the phoneme counting accuracy of the algorithm with that of expert human listeners. Although large errors of up to 11% between the automatic and human counting were scarce and can be regarded as outliers, and the majority of counting errors were much smaller, the algorithm should definitely be improved in order to substantially decrease this maximal error rate.

For the present proof of concept of the entire application, however, only reasonable correlation with the human counting was sought. Moreover, since the application is targeted to practice usage by a single speaker, and since the "slow", "normal" and "fast" speaking rate ranges can be tuned to each user by the therapist, the main requirement of the algorithm is to significantly differentiate for each subject between the "normal" rate range and the "fast" and "slow" ranges. Our experiments indicated that the algorithm was successful in presenting a statistical difference between the three speaking rates in terms of phoneme counting, for eight of the ten subjects.

An analysis of the phoneme counting errors, however, provided some insight into the shortcomings of the current algorithm: The results demonstrate inferior accuracy for slow speech compared to normal and fast speech rate. The algorithm phoneme counting in those cases was higher than the counting by human listeners. One reason for this may be that in slow reading the speakers altered their normal speech rate by increasing the number and duration of pauses between words, and/or elongating speech sounds (LaSalle, 2015; Tiffany, 1980). This tendency was also commented upon by the human listeners who counted the phonemes. Some pauses were filled with extra syllables like "eh", or gasps and breaths. Although the recordings were performed in a quiet room, some ambient noise could be heard in the recording. The pre-processing stage of our algorithm was successful in decreasing this noise but could not eliminate it completely. The large variability in mean speaking rate among subjects may also indicate that "slow reading" was performed differently by the different subjects. The algorithm's phoneme counting was less accurate for fast speech recordings as well, although to a lesser degree compared to the slow speech counting. This could be attributed to the speakers "racing" through the text, skipping or shortening phonemes or in some cases stopping to breathe in the middle of a sentence or word. Both these phenomena may have induced noises that the algorithm was not able to interpret correctly.

The algorithm demonstrated inferior accuracy in the computation of female subjects' speaking rates: over-counting was observed for the female speakers compared to the manual count. This behavior was most distinct at the slow speaking rate. As in other speech processing technologies, gender-related differences may necessitate a different model for male and female speakers (Andrews, Kohler, Campbell, Godfrey, & Hernández-Cordero, 2002; Vergin, Farhat, & O'Shaughnessy, 1996). Additional investigation of the speech processing stages in the algorithm in a larger cohort may yield further insight into the parameters discriminating males and females in terms of speaking rate.

A major challenge in the algorithm development was to set the thresholds for the STM values that determined a phoneme boundary. Our preliminary algorithm uses an adaptive threshold mechanism and demonstrated acceptable accuracy compared to the human phoneme count. The STM algorithm will be further improved in our future studies by additional optimization of the threshold parameters.

The algorithm was tested on read Hebrew speech. The results reported here may not apply to other languages. Indeed, the linguistic characteristics of Hebrew compared to other languages, such as different consonant and vowel durations, may require tuning of the algorithm for each new language, similar to many other speech technologies (Rabiner & Juang, 1993).

Our application was tested on subjects without speech disorders. In order to confirm its suitability for the needs of stutterers, future studies will test the application on speech disorder clinic patients. Further study of both healthy (non-stuttering) individuals and individuals with speech disorders (stuttering) will provide insight into how these individuals can follow the speech rate guidance of the application and whether their speech quality improves as a result.

Summary

This article describes a system that monitors speaking rate, targeted to aid the treatment of patients with speech disfluency. The system includes two modes: a patient application for home practice and a therapist application for monitoring, supervising and storing of the patient's practice. An algorithm for speaking rate computation, based on phoneme counting, is implemented in an Android smartphone application where the results are presented to the user both graphically and numerically in a real-time, user-friendly manner. This application has the potential to enhance speech therapy by providing the patient with timely feedback on his/her practice to control his/her speaking rate. The patient's recordings are stored on a cloud and are uploaded to therapist's website. Consequently, the therapist can send the patient real-time as well as periodical feedback about his/her practice. The speech therapist can thus benefit from the application by having a full record of his/her patients' practice and a tool to supervise and flexibly change exercise schemes. This application may be effective in other speech disorders where an improvement of speech rate is desired. Such cases include controlling a too fast speech, changing the rate of speech for better intelligibility, or maintaining a uniform speaking rate over time. The application may also be useful for practicing speech volume control.

The new tool integrates well in the era of telemedicine. The application will lead to treatment improvement in terms of both efficacy and cost. The practice can be carried out by the patient at home, reducing the frequency of his/her follow-ups in the clinic. The therapist can conveniently follow the patients' progress in real time or off line according to the treatment definitions, as well as receive statistics about a subject or a cohort of patients for clinical analysis as well as research.

This preliminary study will be pursued in order to improve and adapt the STM algorithm and the application to adults, children and different stuttering conditions. Further tests on both controls and patients with speech disorders, as well as improvements of the algorithm accuracy, can make this application an important tool in the speech therapy process, enhancing both the productivity of the process and the quality of treatment.

References

Amir, O., & Levine-Yundof, R. (2013). Listeners' attitude toward people with dysphonia. *Journal of Voice*, *27*, 524 [e521-524. e510].

Andrade, C. R. F. d., Cervone, L. M., & Sassi, F. C. (2003). Relationship between the stuttering severity index and speech rate. *Sao Paulo Medical Journal*, *121*, 81–84.

Andrews, W. D., Kohler, M. A., Campbell, J. P., Godfrey, J. J., & Hernández-Cordero, J. (2002). Gender-dependent phonetic refraction for speaker recognition Acoustics, Speech, and Signal Processing (ICASSP). *2002 IEEE International Conference on* (Vol. 1) [pp. I-149-I-152].

Awad, S. S. (1997). The application of digital speech processing to stuttering therapy. In *Instrumentation and Measurement Technology Conference, 1997. IMTC/97. Proceedings. Sensing, Processing, Networking., IEEE 2* (pp. 1361–1367).

Bland, J. M., & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, *327*, 307–310.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment*, *6*, 284.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, *70*, 213.

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*, 385–390.

Duchin, S. W., & Mysak, E. D. (1987). Disfluency and rate characteristics of young adult, middle-aged, and older males. *Journal of Communication Disorders*, *20*, 245–257.

Dusan, S., & Rabiner, L. (2006). On the relation between maximum spectral transition positions and phone boundaries. In *Ninth International Conference on Spoken Language Processing*.

Fleiss, J. L., Levin, B., & Paik, M. C. (2013). *Statistical methods for rates and proportions*. John Wiley & Sons.

Grayden, D. B., & Scordilis, M. S. (1994). Phonemic segmentation of fluent speech. *Acoustics, Speech, and Signal Processing, 1994. ICASSP-94. 1994 IEEE International Conference on* (Vol. 1) (p. 1) [pp. I/73-I/76 vol. 71].

Hargrave, S., Kalinowski, J., Stuart, A., Armson, J., & Jones, K. (1994). Effect of frequency-altered feedback on stuttering frequency at normal and fast speech rates. *Journal of Speech, Language, and Hearing Research*, *37*, 1313–1319.

Huang, X., Acero, A., Hon, H.-W., & Foreword By-Reddy, R. (2001). *Spoken language processing: A guide to theory, algorithm, and system development*. Prentice Hall PTR.

Hudock, D., Dayalu, V. N., Saltuklaroglu, T., Stuart, A., Zhang, J., & Kalinowski, J. (2011). Stuttering inhibition via visual feedback at normal and fast speech rates. *International Journal of Language & Communication Disorders*, *46*, 169–178.

Ingham, R. J., Kilgo, M., Ingham, J. C., Moglia, R., Belknap, H., & Sanchez, T. (2001). Evaluation of a stuttering treatment based on reduction of short phonation intervals. *Journal of Speech, Language, and Hearing Research*, *44*, 1229–1244.

Kalinowski, J., Armson, J., Stuart, A., & Gracco, V. L. (1993). Effects of alterations in auditory feedback and speech rate on stuttering frequency. *Language and Speech*, *36*, 1–16.

Kalinowski, J., Stuart, A., Sark, S., & Armson, J. (1996). Stuttering amelioration at various auditory feedback delays and speech rates. *International Journal of Language & Communication Disorders*, *31*, 259–269.

Kupryjanow, A., & Czyzewski, A. (2010). Real-time speech-rate modification experiments. In *Audio engineering society convention 128*. Audio Engineering Society.

LaSalle, L. R. (2015). Slow speech rate effects on stuttering preschoolers with disordered phonology. *Clinical Linguistics & Phonetics*, *29*, 354–377.

Lutz, K. C., & Mallard, A. (1986). Disfluencies and rate of speech in young adult nonstutterers. *Journal of Fluency Disorders*, *11*, 307–316.

Maguire, G. A., Yeh, C. Y., & Ito, B. S. (2012). Overview of the diagnosis and treatment of stuttering. *Journal of Experimental & Clinical Medicine*, *4*, 92–97.

Morgan, N., & Fosler-Lussier, E. (1998). Combining multiple estimators of speaking rate. In *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on 2* (pp. 729–732).

Morgan, N., Fosler-Lussier, E., & Mirghafori, N. (1997). Speech recognition using on-line estimation of speaking rate. *Eurospeech*, *Vol. 97*, 2079–2082.

Müller, R., & Büttner, P. (1994). A critical discussion of intraclass correlation coefficients. *Statistics in Medicine*, *13*, 2465–2476.

Obaidat, M. S., Sevillano, J. L., & Filipe, J. (2012). E-business and telecommunications. *International Joint Conference, ICETE 2011* (Vol. 314) [Revised Selected Papers].

Oppenheim, A. V., Schafer, R. W., & Buck, J. R. (1989). . *Discrete-time signal processing* (Vol. 2) Prentice-hall Englewood Cliffs.

Pandharipande, M. A., & Kopparapu, S. K. (2011). Real time speaking rate monitoring system. In *Signal Processing, Communications and Computing (ICSPCC), 2011 IEEE International Conference on* (pp. 1–4).

Pellegrino, F., Farinas, J., & Rouas, J. (2004). Automatic estimation of speaking rate in multilingual spontaneous speech. *Speech Prosody 2004, International Conference*.

Pfau, T., & Ruske, G. (1998). Estimating the speaking rate by vowel detection. *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on* (Vol. 2) (pp. 945–948).

Pfitzinger, H. R., & Itzinger, H. R. P. (1998). Local speech rate as a combination of syllable and phone rate. *The 5th International Conference on Spoken Language Processing*.

Rabiner, L., & Juang, B.-H., (1993). Fundamentals of speech recognition.

Ramus, F., (2002). Acoustic correlates of linguistic rhythm: Perspectives.

Shrawankar, U., & Thakare, V. M., (2013). Techniques for feature extraction in speech recognition system: A comparative study. arXiv preprint arXiv:1305.1145.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428.

Siegler, M., & Stern, R. M. (1995). On the effects of speech rate in large vocabulary speech recognition systems. *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on* (Vol. 1) (pp. 612–615).

Starkweather, C. W. (1987). *Fluency and stuttering*. Prentice-Hall, Inc.

Stuart, A., Kalinowski, J., Rastatter, M. P., Saltuklaroglu, T., & Dayalu, V. (2004). Investigations of the impact of altered auditory feedback in-the-ear devices on the speech of people who stutter: Initial fitting and 4-month follow-up. *International Journal of Language & Communication Disorders*, *39*, 93–113.

Tiffany, W. R. (1980). The effects of syllable structure on diadochokinetic and reading rates. *Journal of Speech, Language, and Hearing Research*, *23*, 894–908.

Vergin, R., Farhat, A., & O'Shaughnessy, D. (1996). Robust gender-dependent acoustic-phonetic modelling in continuous speech recognition based on a new automatic male/female classification. *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on* (Vol. 2) (pp. 1081–1084).

Verhasselt, J. P., & Martens, J.-P. (1996). A fast and reliable rate of speech detector. *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on* (Vol. 4) (pp. 2258–2261).

Wang, D., & Narayanan, S. S. (2007). Robust speech rate estimation for spontaneous speech. *Audio, Speech, and Language Processing, IEEE Transactions on*, *15*, 2190–2201.

Welling, L., & Thomson, L., (2004). PHP and MySQL Web Development (Developer\'s Library).

Wingate, M. E., (1976). Stuttering: Theory and treatment: Irvington.

Xie, Z., & Niyogi, P. (2006). Robust acoustic-based syllable detection. In *INTERSPEECH*. Citeseer.
Ziolko, B., Manandhar, S., & Wilson, R. C. (2006). Phoneme segmentation of speech. *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on* (Vol. 4) (pp. 282–285).

Vered Aharonson is an associate Professor at the School of Electrical Engineering, University of the Witwatersrand, Johannesburg, South Africa. Formerly, prof Aharonson was an associate professor at Afeka, Tel Aviv academic college of Engineering, and founder of the ACLP, Afeka research Center of Language Processing, a member of Humaine – a European Network of Excellence and studied systems to register, model and/or influence human emotional states from speech, and a research fellow at the Eaton Peabody Laboratory, Harvard University. Prof. Aharonson holds an M.Sc. Physics from the Technion, Israel Institute of Technology and Ph.D. Electrical Engineering from Tel Aviv University.