Diagnostic system for speech articulation and speech understanding

A. Czyzewski. B. Kostek, H. Skarzynski
Institute of Physiology and Pathology of Hearing
Warsaw, Poland
e-mail: speech@telewelfare.com

With the increase in access to multimedia computers, speech training can be made available to patients with no continuous assistance required from speech therapists. Another function the system can easily perform is screening testing of speech fluency providing directed information to patients who have various speech disorders and problems with understanding speech. The idea underlying the proposed system is a programmed speech therapy training algorithm consisting of diagnostic tools and rehabilitation devices connected with it. The first function the system has to perform is data acquisition where information about the patient's medical history is collected. This is done through electronic questionnaires. The next function is analysis of the speech signal articulated by the patient when prompted by the computer followed by some multimedia tests carried out in order to assess the subject's ability to understand speech. Next, the results of the electronic questionnaire, the patient's voice and patient's reactions are automatically analyzed. Based on that the system automatically diagnoses possible speech disorders and how strong they are. A large number of schoolchildren was tested with this method. In the paper foundations of applied speech testing method and obtained results will be demonstrated.

I. INTRODUCTION

The aim of this paper is to present a system for people with speech disorders causing the inability to engage in a fluent articulation of speech. With the increase in access to multimedia computers, speech training can be made available to patients with no continuous assistance required from speech therapists. Moreover, as computer hardware becomes more and more miniaturized, some speech training or correction algorithms can now be implemented into small wearable devices that patients can use in everyday life. The system includes some of the new applications of computer technology to speech articulation correction, particularly useful in cases of stuttering [2].

In very rich literature on this subject [1][6]-[9][11][13]-[17] employed solutions do not address the application of multimedia computers or today's communication technology to the problem of speech therapy. On the other hand, the existing computer programs for stutterers do not implement adequate real-time speech processing methods for the correction of disordered speech. Such programs offer various speech exercises for stutterers, generate training material, however, they do not provide any means for a direct electronic analysis of speech or its correction through modification of the auditory feedback loop.

The presented system offers a more extensive and better use of personal computers by providing a fully integrated system of speech therapy employing real-time signal processing and network communication features of contemporary computers.

II. SYSTEM PRESENTATION

The system offers a new type of interactive speech therapy service to people with speech disorders. In particular, the computer software designed according to the system concept, can be installed on local hosts or on computer network stations. The software provides vital assistance to a speech therapist or patient in the area of diagnosis and rehabilitation of speech disorders, especially stuttering. Another function the system can easily perform is providing directed information to patients who have various speech disorders and problems with understanding speech.

The idea underlying the system is a programmed speech therapy training algorithm, which consists of diagnostic tools and rehabilitation devices connected with it. The device is a programmable recorder and player or modifier of speech in the auditory feedback loop, which can be realized on the basis of computer software or as a specialized hardware unit. Nowadays it can be assumed that the potential user of the system has access to a typical multimedia computer with a sound card and a headset. Installed on the computer is a local or network software that meets the below functions. The first function the system has to perform is data acquisition where information about the patient's medical history is collected. This is done through a system of electronic questionnaires. The next function is analysis of the speech signal articulated by the patient when prompted by the computer. The analysis is performed when the patient reads out fragments of a specially selected text displayed on the screen. To implement the function, the computer is provided with speech analysis software [5][10]. As speech is received via a microphone, the software detects errors and articulation defects. Next, the results of the electronic questionnaire and the patient's voice are analyzed. Based on that the system automatically diagnoses any possible speech disorders and how strong they are. The system informs the patient who has been diagnosed with a potential speech disorder that they need to consult a speech therapy or phoniatry specialist to have the diagnosis confirmed. In particular, the patient-specialist contact can be facilitated using teleconference features, which can be integrated into the system. The patient may contact a specialist and have the diagnosis confirmed. Then the patient may look up a therapy program in the system menu to match the disorder. The program recommends the frequency and type of speech articulation or understanding training, which the computer will help to set up. If the speech disorder involves stuttering, the essential part of the therapy is training the articulation using a program for modifying signals in the auditory feedback loop which is between the microphone used by the patient and the earphones. The signals in the loop can be modified by the computer or by external device. The software installed in the computer works in real time and causes a delay of the signal or reverberation or frequency shifting or generates the chorus effect or masking of speech with noise or another sound. As an option, once selected and matched to the patient's needs, the signal modification algorithm settings can be sent from the computer to a miniaturized sound processor. The patient can use the processor together with a small microphone and earphones in everyday life. In the case of patients who have trouble understanding speech, the external electronic device can also receive word material to train speech understanding. The material is a set of lexical units and structures, which the patient should listen to systematically. The miniature external device can also be used to record the patient's speech, which can then be transmitted to the computer for further playback and analysis to assess articulation.

The diagnostic and rehabilitation system significantly assists one in diagnosing speech disorders and problems with speech understanding. This feature is especially important for screening testing. Also, once the disorder has been diagnosed, the system helps to organize a speech therapy training in just a short time. In cases of stuttering, the system ensures a direct access to electronic correction tools, which facilitate training and learning to speak fluently. The system also assists in electronic speech prosthesis fitting [3][4][18]. In particular, the package of services is available via the Internet and on typical personal computers, which makes cost effective dissemination of the service possible. In such a way it creates a new dimension of medical and telemedical speech therapy and phoniatry service.

III.a. Electronic questionnaire

The electronic questionnaire in the system for diagnosing and rehabilitating speech disorders, contains several a dozen questions about the occurrence of basic disorders that impair verbal communication. Preferably, questions about the particular disorders are asked in an indirect manner, meaning that they pertain to everyday situations. Based on that conclusions are drawn on speech articulation and speech understanding problems or problems with writing down a text that is being dictated to the subject. The assumption is not that the patient, who can just as well be a child, has the necessary knowledge to be able to answer specialist questions. In this way, the system makes a preliminary test of how the speech articulation organs and kinesthetic speech control function. As an option, the computer system using a speech synthesis program or pre-recorded sound can dictate fragments of a text which the subject is expected to input into the computer using the keyboard or a speaker-independent speech recognition system. Also the test that involves typing of a text being dictated can be repeated with the image of the person face reading it displayed on the computer screen. Thanks to that it is possible to compare the results of speech understanding with and without lip reading. This can lead to conclusions as to the role of the subject's visual speech center. In this way, the data collected through the electronic questionnaire by the system, can be supplemented with the results of verbal, auditory and auditory and visual tests all of which increases the probability of a correct diagnosis. The data, once logically processed by the computer using an algorithm produced by speech therapy and phoniatry specialists, and following the analysis of the results of the tests which use multimedia techniques, can then help to establish whether the subject falls under a risk group, in particular a group of people who potentially may have difficulty with communicating or have no such problems. In particular, the analysis of the data set acquired in the course of tests, can be supported by decision algorithms from the soft computing area, such as neural networks, fuzzy logic, rough set method or genetic algorithms [12]. Thanks to that the algorithm which learns on the basis of the data sets can automatically recognize the particular patient profiles.

Having recognized the patient profile using the system, the system can then, as an option, refer the subject for consultation with a specialist to have further testing done in order to verify the automatic diagnosis. Next, the patient who has been diagnosed with speech problems is instructed on how to use the therapeutic tools that are available in the system. In particular, for each of the risk groups and each diagnosed speech disorder degree, the system has a different therapy procedure. Based on the procedure speech articulation or speech understanding exercises are programmed. They way it works is that the computer displays a programmed material to be read out and at the same time activates speech analysis software. In the system the material can either be displayed on the computer screen or using a known software based speech synthesis or speech play back. The role of the patient is to read out the displayed text or repeat utterances. The utterances can be received by a microphone connected to the computer sound interface and analyzed by a digital speech processing software.

III.b. Compact Digital Speech Therapy Unit

The system can send data associated with the patient profile to a miniaturized speech therapy device working in real time, namely CDSTU - Compact Digital Speech Therapy Unit. The design of the miniaturized speech therapy device CDSTU and its operation are described on the basis of Fig. 1. The device consists of a microcomputer **1,** which contains a known two-way sound interface **2**. Connected to its input is the microphone from the headset **3**. Connected to its output are the earphones from the same headset. The subject can listen to the material via the earphones which are part of the headset **3** connected with the device **1**. As an option, the device **1** in its internal memory can also record patient speech using the microphone from the headset **3** and send it to the computer **5** via the interface **4** for playback and analysis. The device **1** communicates with the external computer **5** via a typical interface **4**, e.g. a USB (Universal Serial Bus) which can also be used to send the settings regarding the modification of the auditory feedback loop. The system operates in such a way that: the microcomputer device **1** through the two-way serial interface **4** receives data from an external computer **5** which in particular, could be a personal computer equipped with software that enables the sending of speech samples or program settings. The program installed in the CDSTU device **1** after disconnecting this device from the computer is able to processes the speech signal according to received algorithm whose role is to delay the speech signal or generate reverberation or chorus effect or perform frequency shifting up or down the spectrum scale or generate a masking noise at a given loudness level. Optionally, the microcomputer device **1** can receive speech signal samples for playback from computer **5** to which it is connected via the interface **4,** or send to the computer **5** samples of patient speech recorded in the microcomputer device memory **1** using the microphone from the headset **3** in order to carry out analyses employing the software installed on the computer **5** whose objective is to determine whether the recorded and transmitted speech was correctly articulated by the patient.

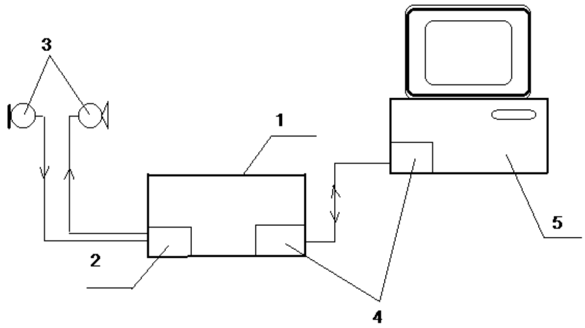

Fig. 1 Block diagram of the speech-therapy system

IV. METHODOLOGY OF AUTOMATIC ANALYSIS OF STUTTERED SPEECH

The presence of vocal tone is a substantial feature of voiced sounds of speech. This tone determines the stimulation of the vocal track. The voicing is caused by vocal cord vibrations. Generally, it may be considered as a sawtooth signal that contains all harmonics. The shape of the vocal track determines the resonances in it, and consequently the timbre of the resulting sound. Moreover, changes of its shape determine the articulation. The vocal tone constitutes the pitch of the speech signal. In stuttered speech, there are spontaneous, sometimes periodic muscle cramps, causing blocking or other disturbances of speech. This may be easily observed by analyzing the speech signal only in the time domain in order to get some data for statistical analysis [14][17]. Analysis of the stuttered speech signal may also be useful for assessing different types of stuttered speech correction systems. In this case, analysis in the frequency domain should be recommended. Cepstral analysis and cepstral smoothing may be applied to detect vowels, and assist in the study of such speech disorders as syllable or vowel repetition and vowel prolongation. It may be said that it is advisable to focus the attention on analyzing the vocal tone in dysfluent speech, especially it seems important to estimate its quantitative changes.

Investigation of the changes in pitch of the vocal tone was based on a modified cepstral analysis, which consists in several steps:

- compression of the dynamics of the spectrum,
- reduction of the spectrum band for cepstral analysis,
- estimation of the frequency of the vocal tone.

These modifications lead to a better discernibility of the cepstral maximum that is the consequence of the presence of vocal tone in the analyzed speech.

In addition, frequencies of formants were computed by interpolating second order polynomials for a cepstral smoothed spectrum. This method searches for local maxima of the smoothed spectrum and computes the value on the frequency axis for which the interpolated function reaches maximum. This method is illustrated in Fig. 2.

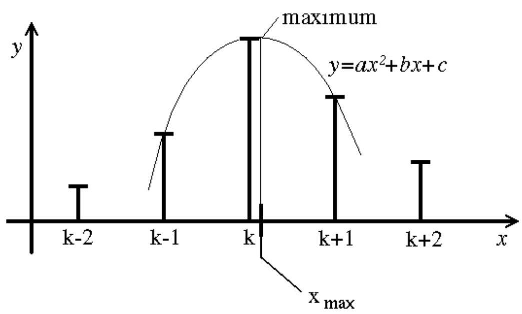

Fig. 2 Graphical presentation of the method of square interpolation for the formant estimation

The general formula for x_{max} is derived from the square function, with the assumption that for $x=k$, the measurement data reach a local maximum. Therefore, using a shift to the left in the coordinate system one can write about k that:

$$x_{\max} = -\frac{b}{2a} = \frac{y(k-1) - y(k+1)}{2(y(k+1) - 2y(k) + y(k-1))} \tag{1}$$

where:
a, b – coefficients at x^2 and x of the polynomial

Due to the shift of the system and considering the resolution of the spectral analysis, one can finally write the formula for the frequency of the analyzed formant:

$$f_{\max} = r_f \cdot \left(k + \frac{y(k-1) - y(k+1)}{2(y(k+1) - 2y(k) + y(k-1))} \right) \tag{2}$$

IV. CONCLUSION

The presented system is designated for people with speech disorders causing the inability to engage in a fluent articulation of speech. With the increase in access to multimedia computers, speech training can be made available to patients with no continuous assistance required from speech therapists. Moreover, as computer hardware becomes more and more miniaturized, some speech training or correction algorithms can now be implemented into small wearable devices that patients can use in everyday life. The system can send data associated with the patient profile to a miniaturized speech therapy device working in real time, namely CDSTU - Compact Digital Speech Therapy Unit. This device is implementing real-time algorithms for the auditory feedback loop modification which are controlled by the personal computer and offers also the possibility of recording, transmitting and playing back the speech material scheduled to the training.

References

[1] Archibald, L. de Nil, L. F. (1999). “The Relationship between stuttering severity and kinesthetic acuity for jaw movements in adults who stutter”, J. Fluency Disorders, 24, 1, 25-42.

[2] Czyzewski, A., Kostek, B., Skarzynski, H. (2000). Training System and Device for the Fluent Speech Articulation, International Patent Application No. PCT/PL00/00081.

[3] Czyzewski, A., Kostek, B. and Roland-Mieszkowski M., (1993). "Digital Speech Aid Based on the Modification of the Auditory Feedback Loop", Presented at the Eurospeech'93, 3rd Eurospeech Congress, Berlin, Germany.

[4] Czyzewski, A., Kostek, B. and Roland-Mieszkowski M. (1994). "Digital Speech Aid to decrease stuttering - clinical results and patient's reaction", First World Congress of Fluency Disorders, Chapter 11, 526-529, Munich, Germany.

[5] Czyzewski, A. and Skorka, P. (1996). "Modification of the Auditory Loop-Effects on Hearing and Speech Production. 100th Audio Engineering Society Convention, preprint No. 4148 (C-1), Copenhagen, Denmark.

[6] Howell, P. and Wingfield, T. (1990). "Perceptual and acoustic evidence for reduced fluency speech in the vicinity of stuttering episodes", Lang. Speech 33, 1, 31-46.

[7] Howell, P., Staveley, A., Sackin, S. and Rustin, L. (1995). "Comparison of speech motor development in stutterers and fluent speakers between 7 and 12 years old", J. Fluency Disorders, 20, 1, 243-255.

[8] Howell, P., Staveley, A., Sackin, S. and Rustin, L. (1998). "Methods of interval selection, presence of noise and their effects on detectability of repetitions and prolongations", J. Acoust. Soc. Amer., 104, 6, 3558-3567.

[9] Howell, P., Au-Yeung, J. and Pilgrim, L. (1999). "Utterance rate and linguistic properties as determinants of lexical dysfluencies in children who stutter", J. Acoust. Soc. Amer., 105, 1, 481-490.

[10] Kaczmarek, A. and Skorka, P. (1997). "Investigation of the Pitch of the Vocal Tone for Stuttered Speech", 102nd Audio Engineering Society Convention, preprint No. 4486 (L7), Munich, Germany.

[11] Kalinowski, J., Armson J. and Stuart A. (1995). "Effect of Normal and Fast Articulatory Rates on Stuttering", J. Fluency Disorders, 20, 3, 293-302, 1995.

[12] Kostek, B. (1999). "Soft Computing in Acoustics, Applications of Neural Networks, Fuzzy Logic and Rough Sets to Musical Acoustics, Studies in Fuzziness and Soft Computing", Physica Verlag, Heilderberg, New York.

[13] Kuniszyk-Jozkowiak, A. (1995). "The statistical analysis of speech envelopes in stutterers and non-stutterers", J. Fluency Disorders, 20, 1, 11-23.

[14] Kuniszyk-Jozkowiak, W. (1996). "A comparison of speech envelopes of stutterers and nonstutterers", J. Acoust. Soc. Amer., 100, 2, 1105-1110.

[15] Michaelis, D., Frohlich, M. and Strube, H. W. (1998). "Selection and combination of acoustic features for the description of pathologic voices", J. Acoust. Soc. Amer., 103, 3, 1628-1639.

[16] Lee B. (1950). "Effects of delayed speech feedback", J. Acoust. Soc. Amer., 22, 824-826.

[17] Robb, M. and Blomgren, M. (1997). "Analysis of F2 Transitions in the Speech of Stutterers and Non-stutterers", J. Fluency Disorders, 22, 1, 1-16.

[18] Roland-Mieszkowski, M., Czyzewski, A. and Kostek, B. (1995). "DSA (Digital Speech Aid) - a new device to decrease or eliminate stuttering. First World Congress of Fluency Disorders", Chapter 11, Munich, 535-539, Munich, Germany.