Voice Controlled Quiz for People with Hearing Impairment

Goran Bujas[1(✉)], Luka Bonetti[2], Zeljka Car[1], and Marin Vukovic[1]

[1] Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, Zagreb, Croatia
{goran.bujas,zeljka.car,marin.vukovic}@fer.hr

[2] Faculty of Education and Rehabilitation Sciences, University of Zagreb, Zagreb, Croatia
luka.bonetti@erf.hr

Abstract. Persons with complex communication needs have difficulties in production and/or understanding of oral and/or written language. Software applications have great potential in speech rehabilitation, especially when providing visual feedback of the produced voice to users thus enabling better rehabilitation and increased user motivation. This paper presents software application for people with hearing impairment that has a form of voice controlled quiz with visual feedback of the users' input voice frequency. The proposed application is evaluated through a case study of using it as an input for on-line application for hearing-aid evaluation.

Keywords: Biofeedback · Hearing impairment · Voice controlled quiz · Complex communication needs

1 Introduction

Complex communication needs is a term referring to significant speech, language, motor and/or cognitive impairments which restrict person's ability to participate independently in society. Persons with complex communication needs are not able to communicate temporarily or permanently by using speech that is the basic mean of communication between people. They also have difficulties in production and/or understanding of oral and/or written language. They are of different ages, abilities, social status, nationality, etc.

Software solutions managed by voice have great potential in speech rehabilitation of people with complex communication needs. Proper selection of design and development technology could result in software applications that will motivate users, especially children, during rehabilitation.

In this paper we present software application for people with hearing impairment that uses biofeedback during rehabilitation. Biofeedback can be tactile or visual and software applications are most suitable for providing visual feedback. In the presented application, we analyse the user's voice frequency in real time and show its level as a bar on smartphone interface. By lowering or raising voice frequency the user selects answers to questions ("Yes", "No" and "Don't know") which can then be used as an

O. Gaggi et al. (Eds.): GOODTECHS 2016, LNICST 195, pp. 166–175, 2017.
DOI: 10.1007/978-3-319-61949-1_18

input to various other applications such as quizzes. This functionality provides visual feedback of the input frequency to users, thus giving them the ability of better control of the produced frequency. Furthermore, answering questions in this manner can be entertaining which motivates the users for practicing their voice frequency output while using the proposed application.

The rest of the paper is organized as follows. Section 2 explains the multidisciplinary research background and an existing need for software applications managed by voice that could be used in rehabilitation of hearing impairment. The rehabilitation science findings that biofeedback significantly affects the production of standard speech and can improve rehabilitation process is explained in Sect. 3 and presents a motivation for developing the presented software solution. Software prototype and used technologies are presented in Sect. 4. Section 5 gives a case-study of using developed prototype with on-line application for hearing-aid evaluation. Finally, we conclude the paper with future research directions in Sect. 6.

2 Research Background

One of the main achievements of EU funded research[1] was establishment of multidisciplinary Competence Network for Innovative Services for Persons with Complex Communication Needs. The Competence Network has been established to provide a framework for its members to cooperate in education, research, development and innovation, joint applications for projects and establishing a dialogue with European and national government bodies and agencies responsible for the development of an inclusive society. Competence network members are University of Zagreb and University of Osijek faculties in the fields of electrical engineering and computing, education and rehabilitation, psychology and graphic arts. Also the members are parental and professional non-governmental associations, polyclinic for consultative health protection of persons with problems of speech communication and several hi-tech and software SMEs.

Scientists and professionals closely collaborate with persons with complex communication needs and their parents and caretakers, collect and analyse their needs and try to find innovative solutions based on information and communication technology and Design for All concept. The multidisciplinary research is focused on the:

- Innovative web and mobile applications for Alternative and Augmentative Communication (AAC)
- Web and Mobile Prototypes for communication and education
- Accessibility analysis of software solutions and emerging mobile devices
- Accessible native User Interfaces (UIs) and AAC service dynamic content adaptation.

[1] ICT Competence Network for Innovative Services for Persons with Complex Communication Needs "(ICT-AAC)" 03/2013 - 03/2015, http://www.ict-aac.hr/projekt/index.php/en/.

The research is under way for more than seven years and there is a portfolio of about 20 web, Android and iOS applications that are free and can be downloaded from Stores or Competence Network Web site.

The multidisciplinary research presented in this paper is part of Competence Network ICT-AAC activities and described research background.

3 Related Work and Benefits of Using Biofeedback

Control of voice production in speech rehabilitation of children with hearing impairment (HI children) is traditionally treated by visual or tactile feedback – i.e. biofeedback treatment [1]. Compromised hearing and auditory feedback significantly affects the production of standard speech. Therefore, speech training of HI children is often facilitated with visual representation of the essential characteristics of speech signals.

Treatments that are based on these foundations have a positive impact on learning of correct speech production [2], and are especially supported by development of software applications allowing graphical representation of speech movements in real time. Such biofeedback applications are attractive and stimulating, making clinical atmosphere more relaxing and accessible to HI children, allow clearer instructions from the therapists, and in general provide more effective training of voice and speech production [3].

Visual presentation of articulators' movement complements to some extent parts of speech signals that are filtered by hearing impairment. This association between the complementary auditory-visual input and speech production has already been noted at the beginning of the last century [4]. According to the acoustic theory of speech production, acoustic patterns of speech signals are received through hearing, processed and organized as internal experience maps [5].

However, they can deviate from the standard if the acoustic patterns of speech are not appropriately adopted, which is exactly the case with hearing impairment. The results of incorrect mapping due to improper auditory input can be various coexisting articulation and prosodic deviations from standard speech [6].

In addition to articulation errors negatively affecting speech intelligibility, voice pitch and intensity control are also important for speech intelligibility, as well as for positive experience in communication with a person with hearing impairment [7]. Visual feedback can provide a HI child with acoustic and physiological information necessary to control the invisible and complicated respiratory-phonation processes in the background of speech prosody [8]. This information would otherwise remain partially or completely unavailable due to the limitations of discrete hearing discriminatory skills that provide auditory feedback and allow the production of speech.

A HI child who learns to speak can compare his/her production efforts with the given model by tracking visual information, which gives objective feedback on the accuracy of their prosody and/or pronunciation [9].

Several studies that have examined the option of applying visual biofeedback in speech training for HI children proved that electropalatography [9, 10] and software technology of voice-games can have a positive effect on training results and duration [11–15].

Specifically in the case of prosodic skills, it has been shown that computer games that provide visual feedback about acoustic characteristics of phonation can help HI children develop more standard vocal skills, i.e. to achieve better coordination of breathing and phonation, better control of the voice pitch and its variation in speech, better control of intonation, speech intensity, voiced-voiceless contrast, as well as better control of accent and speech tempo. Therefore, it is not surprising that the market already recognizes several biofeedback software solutions, especially for the training of voice skills – a segment of speech production most suitable for their implementation. However, the most important advantage of these biofeedback solutions in clinical sense is in their creativity and innovation, user friendliness and high level of personalization. It is therefore appropriate to continuously offer new software solutions, which can be extended with improvements in speech signal processing as well as graphic and acoustic improvements. This was the goal behind creation of the presented Voice Controlled Quiz.

During prototype development the existing voice management solutions were analysed: Pah! game [16] in which the course of the game depends on the volume of the input sound, therefore it can provide good practice for controlling the voice intensity. Although it can be used in rehabilitation, this game was not primary developed as rehabilitation application.

Magic Voice [17] was created by certified speech and language pathologists to promote improved speech and language. It allows users, primary children, to initiate and move animated scenes using the power of their voice. The application is developed only for iOS devices.

Tiga Talk [18] is speech rehabilitation game and help children to learn how to make 23 core phonetic sounds through voice-controlled games that can improve speech clarity and articulation. It is developed for iOS devices only.

4 Voice Controlled Quiz

In this section we present an application designed for Android smartphones that is controlled by user voice frequency.

The Quiz is designed as a biofeedback tool for vocal training of individuals with hearing impairments. It is implemented as an Android application to be used with smartphones and tablet computers. It features interactive graphics that provide feedback to the users and is controlled by continuous phonation into the device microphone. The phonation is adapted according to the levels of desired pitch, controlled and overviewed by a speech rehabilitation therapist. The application recognizes the desired pitch levels and transforms them into visual representation, thus providing feedback to the user in real time. As the main idea of the bio-feedback is to provide a visual change on the device screen as the results of the change in phonation, such graphic information can be used to support the training of coordination and exhalation, phonation, average pitch and even intensity of voice and their changes in speech.

The idea of the application is to use voice frequency for answering any type of questions with simple answers – “yes”, “no” and “don’t know”. Frequency in lower range produced by the user will select “yes” as an answer, middle range frequency will result in “don’t know” while frequency in high range will be considered as a “no”. The frequency ranges can be configured in terms of both central frequency and the upper and lower frequency definition (range of each answer). Configuring central frequency moves all the required answer ranges along the complete frequency range, thus giving the users the possibility to personalize the application to their basic voice frequency. Without this option, the application would be applicable only for certain groups of users with same or similar voice frequencies. Furthermore, definition of range boundaries, i.e. upper and lower boundaries for each answer, gives the users the ability to practice more generally (using larger ranges) and to fine-tune and focus on the desired frequencies (using smaller ranges).

The application uses TarsosDSP library [19] to perform pitch detection. TarsosDSP is a Java library for audio processing that implements several pitch estimation algorithms. The application enables selection of the following parameters:

- pitch estimation algorithm (used for evaluation)
- the lowest pitch value to be recognized
- the highest pitch value to be recognized
- ranges between each answer
- timeout, defining how long should the user hold a certain pitch in order to be selected as an answer within certain range

Pitch estimation algorithm estimates user’s pitch represented in Hertz (Hz). As each person has different frequency range, application has to adapt to particular user by setting lowest and highest pitch. Users can choose between the following algorithms available through TarsosDSP library:

- Fast YIN algorithm (FFT_YIN) implemented by Matthias Mauch, Queen Mary University, London.
- YIN algorithm described in [20]
- McLeod Pitch Method algorithm (MPM) described in paper [21]
- Dynamic wavelet algorithm described in [22]
- Average Magnitude Difference algorithm (AMDF) implemented by Eder Souza and adapted for particular need of TarsosDPS library.

For evaluation purposes, we used Fast YIN which proved to be acceptable for the given purpose. However, in further evaluation we plan to examine the differences in performance between the available algorithms to see whether another algorithm will yield better recognition than the Fast YIN.

Figure 1 shows application interface. Selection scale is on the left with answers indicated by colours. Timeout gauge is in the middle and boxes indicating selected answers are on the right of the interface. Interface also contains text area with information about user input values and calculated pitch ranges for particular answer, on the right hand side of Fig. 1. This was added for evaluation purposes in order to detect how well the pitch is recognized.

Selection scale represents current user pitch. It is divided into seven areas, three of them (coloured in green, grey and red) representing accepted pitch values for particular answer. Accepted pitch values for particular answer are calculated depending on user inputted values for low and high pitch. The bottom edge of the selection scale has low

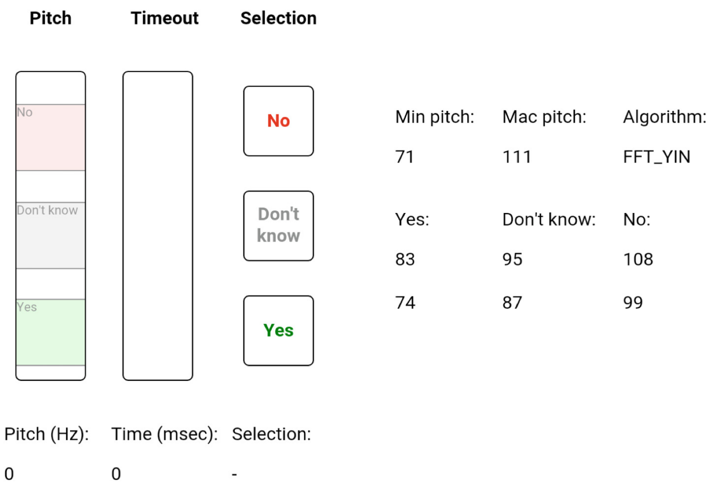

Fig. 1. Application interface

pitch value selected by the user. Top edge of the selection scale has high pitch value selected by the user. Accepted pitch values for particular answer are calculated within these limits.

Timeout scale in the middle represents amount of time necessary for selection to take effect. When user levels his voice within the limits of accepted pitch values for particular answer, the time starts to countdown. After the timeout expires, the answer is selected and is marked so by highlighting the answer in the Selection column.

As the user produces a sound, current pitch value is marked on the pitch scale. Text area on the right with reflects currently measured pitch value. As the user pitch oscillates so does the marker on the selection scale. User has to keep his pitch within the limits of accepted pitch values for particular answer for the amount of time necessary for selection to take effect, defined by timeout.

If the user pitch falls under acceptable levels or goes above acceptable levels the countdown restarts. To be able to select one of the answers a user has to retain his pitch level within the limits of accepted pitch values for particular answer for a defined amount of time (200 ms for evaluation purposes). After countdown time expires selection is made and answer box changes accordingly. An example of positive answer is presented on Fig. 2, with indicators of current pitch and timeout (blue bars).

By using timeout and graphical representation of the user's input pitch in the presented application, we encourage users to try and hold a certain pitch for a defined period of time. This is important as it encourages users to produce a constant pitch for a period of time, which is an important feature in speech production. By gradually narrowing answers' frequency ranges and extending timeout for answer selection, the users can become more and more focused on producing a desired frequency necessary for better speech performance.

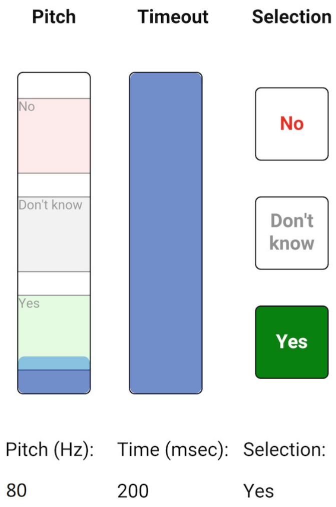

Fig. 2. Positive answer selection and indication

5 Case Study: Using Voice Controlled Quiz with Application for Hearing-Aid Evaluation

As a case study, we used the voice-controlled quiz as an input for existing application used for hearing-aid evaluation. The hearing-aid evaluation application is a web application that was developed previously at the Department of Telecommunications in cooperation with Faculty of Education and Rehabilitation Sciences and is used for user evaluation of various hearing aids.

In short, the users are presented with a pre-recorded audio stories about various every-day topics (e.g. newspaper articles). The stories are grouped into three groups with regards to complexity of the phrases and sentences from the hearing perspective. Furthermore, a background noise is added to the pre-recorded story, with a certain variable percentage of loudness over the original story. There are several available background noise sounds from every-day, such as traffic, people talking, rain and similar. After the users listen to a story with added background noise, they are required to answer several selected questions regarding the story. The questions are presented in a form of a claim and the users have to select whether the claim is true (“yes”), false (“no”) or state that they are not sure (“don’t know”). The questions are used to evaluate how well did the users hear and generally understood the story despite the background noise. This can be used to evaluate the performance of hearing-aid the users are carrying and also track the users’ performance over a period of time. Home screen of the hearing-aid evaluation application is shown on Fig. 3, where the users can select story difficulty level, background noise type and level, as well as the form of questions to show - text, voice or both.

Voice controlled quiz was used as an input for this application, enabling the users to respond to the presented questions by using their voice. This made answering more

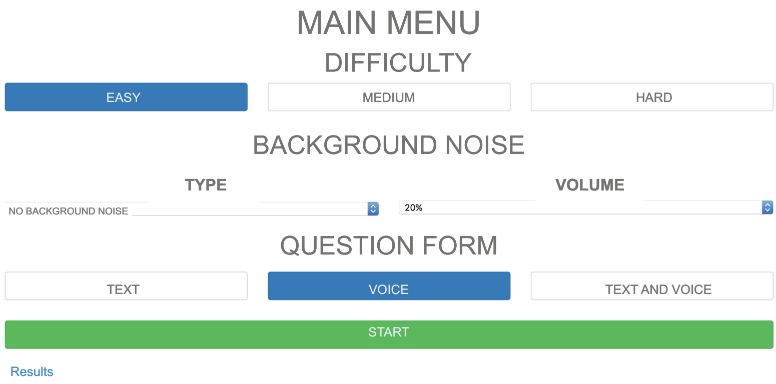

Fig. 3. Hearing-aid evaluation web application

entertaining to the users and encouraged them to focus more on their performance. For this purpose, the voice-controlled quiz application had to be adapted in a way that the user's selected answers triggered requests that were sent to the hearing-aid evaluation web application.

Initial evaluation was performed internally by ten users, mainly developers, students and faculty staff involved in the design and implementation of the application. Each answer that resulted from pitch was considered "selected" after a timeout expired, as explained in Sect. 4. Correct or incorrect selection was reported by the users, i.e. the users told us whether they wanted to select the answer or did the application failed to recognize the correct frequency range they wanted to achieve. In total, the users had to answer 15 questions in order to complete one evaluation iteration, which was mandated by the existing application for hearing–aid evaluation.

The initial results are promising since high majority of the users managed to select the answers successfully, with success rate at over 90% of the answers across all users. Furthermore, the use of voice controlled application gave the users an additional channel for answering questions and made the process more entertaining. Besides hearing-aid evaluation, the users got the ability to practice pitch production which may prove to be beneficial in the long term. However, further evaluation with external users of various backgrounds is required and will be performed in future work. We will observe two groups of users over a longer period of time and track their performance. The first group will be using standard click-based selection of correct answer and the second group will use Voice-controlled quiz application. We expect that the second group will be more motivated for using both applications and may gain benefits from practicing the desired pitch production.

6 Conclusion and Future Work

This paper presents a smartphone application that measures users' input voice frequency and displays the value in form of a bar. This is an example of visual feedback, a subset of biofeedback method that aims to provide better rehabilitation to people with complex communication needs, specifically hearing impairment. The proposed application is used for answering questions and a case study is presented where the application is used for answering questions within an online application for hearing-aid evaluation.

Based on the previous research and related work we can conclude that software applications are suitable for biofeedback method, especially for visual feedback as they can provide feedback in real-time and in a more entertaining manner. Scientific field of software applications for rehabilitation purposes that are managed by user voice is very interesting and promising area, but also very demanding. Each user is different and there is a high need for personalization in various aspects, from user interface, to level of entertainment expected by the user, depending on age.

Nevertheless, technology and advanced sound processing is at a stage where rather complex analysis can be performed even from a web browser and in real time, which will enable more and more sophisticated solutions for the purpose of speech rehabilitation.

Our future work will be focused at analysing different algorithms for sound processing in order to detect whether some users might yield better results when using different algorithms. Furthermore, we will examine the possibilities of applying machine learning methods in order to achieve automated personalization of voice controlled solutions and reduce the need to manually select the recognition boundaries.

Acknowledgement. This work has been fully supported by Croatian Science Foundation under the project 8065 “Human-centric Communications in Smart Networks”.

References

1. Ling, D.L.: Speech and the Hearing-Impaired Child: Theory and Practice. Alexander Graham Bell Association for the Deaf, Washington (1976)
2. Öster, A.-M.: Clinical applications of computer-based speech training for children with hearing impairment. Department of Speech, Music and Hearing, KTH Stockholm, Stockholm (1996)
3. Tye-Murray, N.: Foundations of Aural Rehabilitation: Children, Adults, and Their Family Members, 2nd edn. Thompson Learning - Delmar Learning, Singular Publishing Group, New York (2004)
4. Liberman, A.M., Cooper, F.S., Shankweiler, D.P., Studdert-Kennedy, M.: Perception of the speech code. Psychol. Rev. **74**, 431–461 (1967)
5. Stevens, K.N.: Toward a model for lexical access based on acoustic landmarks and distinctive features. J. Acoust. Soc. Am. **111**, 1872–1891 (2002)
6. Bonetti, L.: Prediktori razumljivosti govora osoba s oštećenjem sluha. Doctoral thesis, University of Zagreb (2008)
7. Abberton, E.: Voice quality of deaf speakers. In: Kent, R.D., Ball, M.J. (eds.) Voice quality measurement, Singular Publishing Group, San Diego (2000)
8. Crawford, E.E.: Acoustic signals as visual biofeedback in the speech training of hearing impaired children, Department of Communication Disorders, University of Cantebury (2007)

9. Dagenais, P.A., Citz-Crosby, P., Fletcher, S.G., McCutcheon, M.J.: Comparing abilities of children with profound hearing impairments to learn consonants using electropalatography or traditional aural-oral techniques. J. Speech Hear. Res. **37**, 687–699 (1994)
10. Pantelemidou, V., Herman, R., Thomas, J.: Efficacy of speech intervention using electropalatography with a cochlear implant user. Clin. Linguist. Phonetics **17**(4–5), 383–392 (2003)
11. Ertmer, D.J., Maki, J.E.: A comparison of speech training methods with deaf adolescents: Spectrographic versus noninstrumental instruction. J. Speech Lang. Hear. Res. **43**(6), 1509–1523 (2000)
12. Massaro, W., Light, J.: Using visible speech for training perception and production of speech for hard of hearing individuals. Department of Psychology, University of California, Santa Cruz, CA 95064 USA (2004)
13. Ryalls, J., Le Dorze, G., Boulanger, H., Laroche, B.: Speech therapy for lowering vocal fundamental frequency in two adolescents with hearing impairments: a comparison with and without SpeechViewer. Volta Rev. **97**(4), 243–250 (1995)
14. Pratt, S.R., Heintzelmann, A.T., Deming, S.E.: The efficacy of using the IBM SpeechViewer vowel accuracy module to treat young children with hearing impairment. J. Speech Hear. Res. **36**(5), 1063–1074 (1993)
15. Shuster, L.I., Ruscello, D.M., Smith, K.D.: Evoking [r] using visual feedback. Am. J. Speech-Lang. Pathol. **1**(3), 29–34 (1992)
16. Taguri, Y.: About PAH! http://ahhhpah.com/. Accessed July 2016
17. Magic Voice. http://pocketslp.com/our-apps/magic-voice/. Accessed July 2016
18. Tiga Talk Speech Therapy Games. http://tigatalk.com/app/. Accessed July 2016
19. TarsosDSP library, https://github.com/JorenSix/TarsosDSP. Accessed July 2016
20. de Cheveigné, A., Kawahara, H.: YIN, a fundamental frequency estimator for speech and music. J. Acoust. Soc. Amer. (JASA) **111**(4), 1917–1930 (2002)
21. McLeod, P., Wyvill, G.: A smarter way to find pitch. In: Proceedings of the International Computer Music Conference, Barcelona, Spain, pp. 138–141, 5–9 September 2005
22. Larson, E.; Maddox, R.: Real-time time-domain pitch tracking using wavelets. In: Proceedings of the University of Illinois at Urbana Champaign Research Experience for Undergraduates Program, Champaign, IL, USA, 3 August 2005