An interactive toolset for speech therapy

André Grossinho, Sofia Cavaco, João Magalhães
CITI, Departamento de Informática
Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa
2829-516 Caparica, Portugal
a.grossinho@campus.fct.unl.pt, {scavaco, jm.magalhaes}@fct.unl.pt

ABSTRACT

This paper proposes a novel approach to include biofeedback in speech and language therapy by providing the patient with a visual self-monitoring of his/her performance combined with a reward mechanism in an entertainment environment. We propose a toolset that includes an in-session interactive environment to be used during the therapy sessions. This in-session environment provides instantaneous biofeedback and assists the therapist during the session with rewards for the patient's good performance. It also allows to make audio-visual recordings and annotations of the session for later analysis. The toolset also provides an off-line multimedia application for post-session analysis where the session audio-visual recordings can be examined through browsing, searching, and visualization techniques to plan the future session.

Author Keywords

Interactive environment; speech therapy; natural interaction; facial analysis; auditory processing; phoneme recognition.

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation (e.g. HCI): User Interfaces; K.3.1. Computers and Education: Computer-assisted instruction (CAI)

INTRODUCTION

Humans start learning how to formulate speech in early childhood. This process appears spontaneously as we gain the perception of speech through hearing, which, combined with lip-reading, gives clues on how to articulate the sounds. The normal development of speech relies on a normal development of the physical and intellectual capacities, as well as on a stimulating environment. Problems in this development may slow down the acquisition of normal speech or lead to speech disorders.

Patients with speech disorders, typically attend speech therapy sessions. In order to have productive sessions, therapists must keep their patients attention and motivation. For this they must use interesting and challenging exercises, which are adapted to the patients abilities and which are somehow rewarded.

The classical therapy usually consists of repetitive exercises with verbal feedback from the therapist. While an adult patient of speech therapy can understand the need for assisted therapy and self-motivate with the idea of the results, usually that is not true for a child that bores and loses motivation easily. Every parent knows that keeping a child's attention is no easy task, more so if the task at hand is enforced such as speech therapy.

Computer training exercises can play a major role in the field of motivation while keeping a child focused. This is even more so if the exercises are actually a game adapted to the child's age and needs. With a captivating and entertaining interface the perspective of therapy can improve along with a kid's motivation. With game like exercises and a visual reward (such a score or an animated measure of performance) a child can actually gain the interest to surpass his last productions of speech, making this a whole new fun and motivating type of task.

Additionally, therapists use biofeedback mechanisms such as mirrors and microphones. Using mirrors allows patients to watch themselves as they pronounce the sound or do a facial exercise, so that they better associate their intention to their actions. Microphones, on the other hand, provide relevant information to the therapist for diagnosis of the patient's conditions and to monitor the patient's progress.

Having that in mind, here we propose a toolset for speech therapy to assist the therapist during and after the sessions. This includes a game-like interactive environment, which has a reward system, to be used during the sessions. While this environment was designed for children therapy sessions, it can easily be adapted for adult therapy. The toolset also includes an off-line multimodal tool for monitoring the progress of the patients and adapting future sessions to the specific needs of the patient.

Due to the entertainment factor of games, therapy may be more effective with the help of such tools. The aim of the proposed game is to help making the therapy sessions more entertaining as to increase the motivation of the patient on doing the therapy exercises. Therefore contributing to more fruitful sessions. Serious games are a useful and effective tool in several types of therapy. The players may experience

ACE'14, November 11–14, 2014, Funchal, Madeira.

http://dx.doi.org/10.1145/2663806.2663869

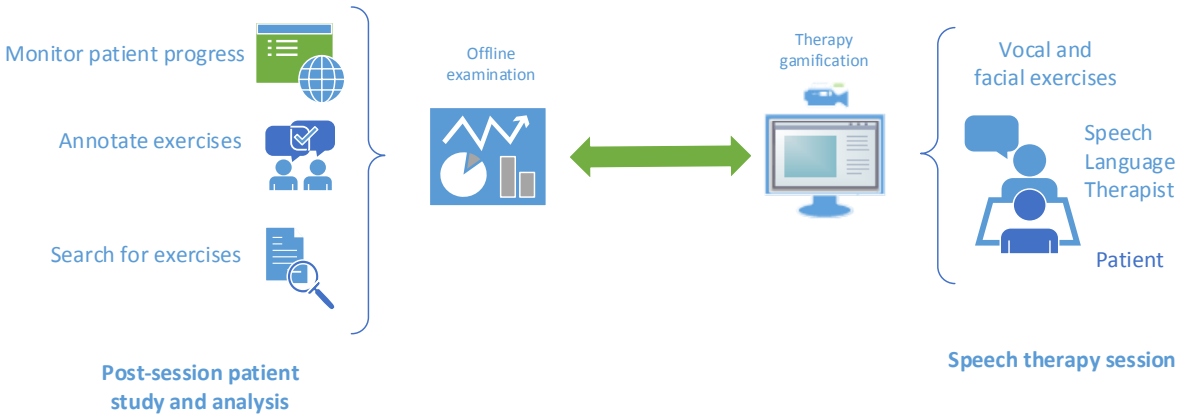


Figure 1. The speech therapy toolset architecture.

more motivation on performing the needed tasks as opposed to when only traditional therapy techniques are used.

Several serious games and computer applications for speech therapy have been proposed [4, 6, 9, 10, 11, 12]. As an example, there is the intelligent system proposed by Bunnell et al., which monitors the progress and suggests exercises for children's therapy sessions for articulation problems [3]. The system uses a game that helps children utter first syllables, then words and latter phrases through a series of tailored difficulty increasing speech production tasks. Another example is the VITHEA system for Portuguese proposed by Abad et al., which has a virtual tutor that presents word naming exercises for the treatment of aphasia [1]. Yet, most existing applications use only audio information.

Here we propose to use both audio and visual information to make the therapy sessions more motivating and more effective. The advantages of combining visual and audio data for speech recognition in speech therapy have been demonstrated by Engwall et al. [5]. They merge audio and video analysis for speech recognition by combining the analysis of facial and vocal tract movements with audio to detect mispronunciation errors. Zhao et al. also combine visual and audio data for speech recognition [13]. In addition, they use visual data for feedback. They use 3D animations of the vocal tract to show the patient the articulated and correct positioning while describing textually the error and the needed correction. OLP (ortho-logo-paedia) is an EU project in speech therapy for different types of pathologies that also uses visual information [2]. The OLP system contains a remote tutoring and monitoring system and allows the visualization of the vocal tract through acoustic-to-articulation mapping in 2D or 3D. Yet, the feedback could be given in a more engaging way, which can be obtained with a game-like environment and reward system as the one proposed here.

Building on our previous work by Mourão et al., the proposed interactive environment combines a reward system with visual stimuli [7, 8]. It substitutes the conventional mirror with the image of the patient in the screen and provides interactive features that contribute to creating an environment where patients can enjoy better their speech therapy exercises. The fact that the tool gives patients a reward for their performance, motivates them to mimic vocal and facial exercises. This contributes to practicing the production of speech sounds, but also to strengthening the muscles of the face and the mouth.

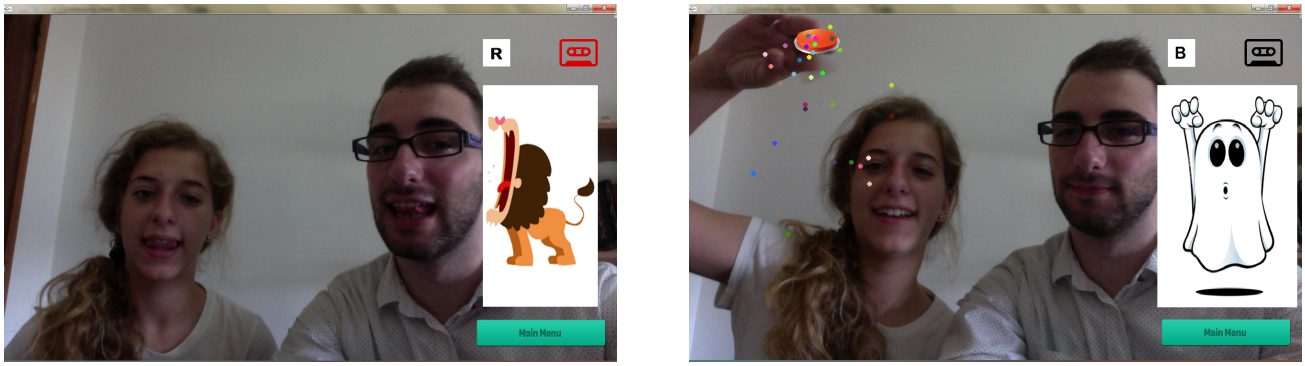
The second part of the framework consists of an off-line multimodal tool for browsing, annotating, searching and visualizing. This gives the therapist access to multimodal information extracted from the current and previous sessions. More specifically, the toolset includes an application for post-session analysis where a patient's audio-visual recordings can be examined and annotated through browsing and searching. It allows monitoring the patient progress and assist in planning future therapy sessions.

THE TOOLSET

The proposed interactive toolset aims at making therapy sessions for speech and language disorders as well as for orofacial myofunctional disorders (a) more motivating and (b) more effective. The aim of the toolset is not to substitute the speech and language therapist (SLT), on the contrary its goal is to help the SLT on planning and executing motivating and fruitful sessions as well as to keep the patients engaged in the therapy exercises. The framework provides the SLT with a comprehensive set of tools that assist him/her in the therapy session and in the post-session study and analysis of the patient status and progress. The toolset includes an interactive environment to be used in-session and an off-line multimodal tool (figure 1). These will be explored in more detail below.

The in-session interactive environment

Traditional behavioral treatment approaches for speech and oral motor disorders, such as facial paralysis can benefit greatly from visual biofeedback for patient motivation purposes, for tracking orofacial motion and for evaluation of the patients speech productions. Existing biofeedback mechanisms such as mirrors are commonly used in speech therapy. These instruments allow patients to watch themselves as they pronounce sounds or do an orofacial movement, which is useful since this way patients understand better the relation between their intentions and their actions. Here we propose to include visual biofeedback in an interactive environment to



(a) (b)

Figure 2. The in-session interactive environment. (a) The interface. (b) The reward button.

provide the patient with a visual self-monitoring of his/her performance in an entertaining and motivating way.

The interactive environment was designed to be used during the therapy sessions (right side of figure 1) and provides a complement to existing biofeedback mechanisms that has the advantage of being more motivating. The environment provides visual stimuli that replaces the mirror. As shown in figure 2.a, the patients are able to see themselves and the SLT in the screen.

In more detail, it consists of a simple serious game that is proposed as a key instrument to keep the patient engaged and reactive to the instant biofeedback provided. During the game the patient is supposed to mimic vocal and facial exercises that are chosen and explained by the SLT and complemented with images provided by the tool. These vocal and facial exercises are an extension of exercises that are usually performed with traditional methods to improve speech sounds production and strengthen the orofacial muscles.

The images provided are associated to phonemes or onomatopoeias and aim at motivating the child to say particular phonemes. For instance, if the SLT wants to exercise the sounds *rr* and *buuh* the chosen images will be a lion and a ghost (see figure 2.a and b). Figure 3 illustrate a few basic examples of images and orofacial expressions associated to phonemes. This way, the interactive environment enables the gamification of the speech therapy session; it creates a healthy competitive environment in which the child can better enjoy his/her speech therapy exercises and allows capturing visual and audio data from the therapy sessions for later analysis.

In addition, the game has a child appropriate reward system that can be activated to reward the child for his/her good performance. This allows the child to interact with the game in a fun way: the child can trigger a virtual button that shows up in the screen and that drops little colorful balls (figure 2.b).

The possibility of exercising in a game environment with biofeedback and a reward system, gives the child more motivation to perform the therapy activities than when only traditional therapy techniques are used. While it is up to the SLT to decide when the child deserves to have the reward of interacting with the virtual button, we are investigating ways of

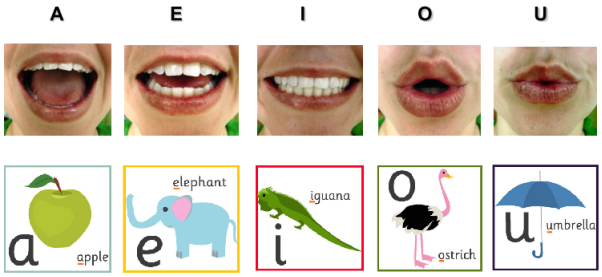


Figure 3. Examples of orofacial expressions and images associated to phonemes (Copyright © 2014 Boquinhas Aprendizagem e Assessoria).

automatically suggesting when to use the button based on the child's performance. In response, the SLT can then decide whether he/she accepts the suggestion or not.

This environment has the capability of recording the images and sounds produced by the patient. To facilitate searching for recorded data, it associates annotations to the recorded data. While in the current version these annotations are still done by the therapist, the proposed game will include automatically generated annotations provided by a phoneme recognition module and an orofacial expression recognition module.

The off-line multimodal tool

The second part of the framework is an off-line multimodal tool for browsing, annotating and searching (left side of figure 1). This tool allows to search and browse the recorded data by name of the patient, phonemes, and other annotations introduced by the SLT, and then watch and listen to the patient speech productions captured during therapy sessions.

While the tool's current version includes only the possibility of watching the captured speech productions, we are currently exploring other more informative visualization options. These include the visualization of spectra, spectrograms, formants and other audio analysis functions. In addition, the tool will include the option of visualizing more than one production simultaneously in order to facilitate monitoring a patients progress as well as to perform cross-patient analysis to help the therapist in finding related cases.

CONCLUSION

This paper proposes an interactive environment that aims at making speech and language therapy sessions more entertaining and consequently more productive. The main novelty of this work is the combination of visual biofeedback and the gamification of the speech therapy. While the success of this approach has been proved in non-therapy areas [7], here we tackle the speech therapy area.

The in-session interactive environment provides the patient with visual self-monitoring of his/her performance combined with a reward mechanism. We have been in close contact with several SLTs, who have tried the tool. They unanimously have given us very positive feedback on this environment and confirmed that it will be a very useful alternative to the mirror.

Besides, the SLTs were particularly interested in the off-line multimodal tool, which allows them to monitor the patient's performance and plan future sessions in more productive ways than those traditionally used. The off-line multimodal tool accesses audio-visual information captured during the therapy sessions and assist the therapist on analyzing that data to evaluate the patient's speech progress, and to measure the coordination of the patient facial muscles in terms of orofacial motion. A patient's audio-visual recordings can be examined through browsing, searching, and visualization features.

Future work

There are many types of feedback that can be presented in a computer based speech therapy tool [9]. A future version of the proposed interactive environment will provide feedback with benefits beyond motivation. Navigational feedback on how to move/position vocal tract articulators (internal perspective) can be given to help achieve a correct production of speech sounds. The measure of performance of a production can help the child compare productions and understand the path to a correct speech production. All the feedback will be delivered in real time and by being non-verbal it will avoid subjectivity.

The in-session environment is proposed as a key instrument to keep the patient engaged and reactive to the instant biofeedback provided by the game. While the proposed tool still depends on the therapist to judge the correctness of the speech pronunciation, in a future version it will perform phoneme recognition, face detection and orofacial expression recognition for this purpose. The latter will build on our previous work by Mourão et al. on facial expression recognition [7, 8].

The offline tool is being extended with features that will allow to make more complete annotation (such as related to the improvements or specific pathologies), more complex searches (such as by similar patients, productions and pathologies) and to visualize the data in several useful ways (eg. spectrograms). These features will help planning future therapy sessions and monitor the patient's progress.

REFERENCES

1. Abad, A., Pompili, A., Costa, A., Trancoso, I., Fonseca, J., Leal, G., Farrajota, L., and Martins, I. P. Automatic word naming recognition for an on-line aphasia treatment system. *Computer Speech & Language* (2012), 1235–1248.
2. AM, P. A. Ö., House, D., and Hatzis, A. Presentation of a new EU project for speech therapy: OLP (ortho-logo-paedia). In *Proceedings of the XV Swedish Phonetics Conference (Fonetik 2002)*, TMH-QPSR (2002).
3. Bunnell, H. T., Yarrington, D., and Polikoff, J. B. STAR: articulation training for young children. In *INTERSPEECH* (2000), 85–88.
4. Danubianu, M., Pentiuc, S.-G., Schipor, O. A., Nestor, M., Ungureanu, I., and Schipor, D. M. TERAPERS-intelligent solution for personalized therapy of speech disorders. *International Journal On Advances in Life Sciences 1*, 1 (2009), 26–35.
5. Engwall, O., Bälter, O., Öster, A.-M., and Kjellström, H. Designing the user interface of the computer-based speech training system artur based on early user tests. *Behaviour & Information Technology 25*, 4 (2006), 353–365.
6. Lopes, C., Veiga, A., and Perdigão, F. A European Portuguese children speech database for computer aided speech therapy. In *Computational Processing of the Portuguese Language*, H. Caseli, A. Villavicencio, A. Teixeira, and F. Perdigão, Eds., vol. 7243 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2012, 368–374.
7. Mourão, A., and Magalhães, J. Competitive affective gaming: winning with a smile. In *ACM Multimedia* (2013), 83–92.
8. Mourão, A., Borges, P., Correia, N., and Magalhães, J. Facial expression recognition by sparse reconstruction with robust features. In *ICIAR* (2013), 107–115.
9. Öster, A.-M. *Computer-based speech therapy using visual feedback with focus on children with profound hearing impairments*. PhD thesis, KTH, 2006.
10. Rivas, E. Q., and Molina, E. S. A proposal for a virtual world that supports theraphy of dyslalia. In *Proceedings of the 6th Euro American Conference on Telematics and Information Systems*, EATIS '12, ACM (2012), 371–374.
11. Schipor, O. A., Pentiuc, S. G., and Schipor, M. D. Improving computer based speech therapy using a fuzzy expert system. *Computing and Informatics 29*, 2 (2012), 303–318.
12. Tewari, A., Goyal, N., Chan, M. K., Yau, T., Canny, J., and Schroeder, U. Spring: Speech and pronunciation improvement through games, for hispanic children. In *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development*, ICTD '10, ACM (2010), 47:1–47:11.
13. Zhao, G., Barnard, M., , and Pietikainen, M. Lipreading with local spatiotemporal descriptors. *IEEE Transactions on Multimedia 11* (2009), 1254–1265.