A voice-controlled serious game for the sustained vowel exercise

Marta Lopes, João Magalhães, Sofia Cavaco
NOVA LINCS, Departamento de Informática
Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa
2829-516 Caparica, Portugal
mds.lopes@campus.fct.unl.pt, {jmag, scavaco}@fct.unl.pt

ABSTRACT

Speech is the main form of human communication. Thus it is important to detect and treat speech sound disorders as early as possible during childhood. When children need to attend speech therapy it is critical to keep them motivated on doing the therapy exercises.

Software systems for speech therapy can be a useful tool to keep the child interested in keep practicing the therapy exercises. Several software systems have been developed to assist speech and language therapists during the therapy sessions. However most software focus on articulation disorders while voice disorders have been mostly neglected.

Here we propose a voice-controlled serious computer game for the sustained vowel exercise, which is an exercise commonly used in speech therapy to treat voice disorders. The main novelty of this application is the combination of real time speech processing, with the gamification of the speech therapy exercises and the parameterization of the difficulty level.

Author Keywords

Educational Games; Sound Analysis; Voice Disorders; Interactive Environment, Adaptive Difficulty Levels, Real Time Speech Processing

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

The speech learning process starts in early childhood, by hearing and observing the articulatory gestures of speakers [8]. It is with this process that we gain the perception of speech and learn how to articulate sounds. Problems during the speech learning process can cause disorders, such as perception and production difficulties, as well as problems in the phonological representation of sounds. Thus, it is of utmost importance that speech sound disorders (SSD) are detected and treated as early as possible.

While attending speech therapy is very important for children with SSDs, children may loose interest if the exercises are monotonous. Thus, speech and language therapists (SLT) try to make the sessions more interesting, using board games, memory games and book exercises that motivate the children to overcome challenges to win scores and rewards.

Over the years, several software systems were created to assist SLTs during the sessions. Some of these systems focus on the use of computer games to captivate the interest of children, such as sPeAK-MAN [10], Flappy Voice [3], Talker [9], Speech Adventure [7], among others. However, there are systems, such as Articulation Station [4] and Falar a Brincar [11] that do not have sound analysis, which means that they can not automatically validate the exercises, and the presence of an adult is always required to do the validation.

Most of the developed systems focus on articulation exercises, like the set of applications from the LittleBeeSpeech [4], sPeAK-MAN and Star [2], whose main goal is to teach how to correctly produce phonemes, syllables, specific words or even sentences. Others were created for other SSD, like apraxia [3, 5], afasia [1] and cleft lip [7]. On the other hand, voice disorders have been mostly neglected.

Voice exercises are very important to correct voice problems. These problems may appear when we speak in a incorrect voice tone that could damage the vocal cords, which may lead to nodules and cysts, causing hoarseness and in extreme cases, aphonia. One such exercise is the sustained vowel exercise, which SLTs use with children who have voice problems. The goal of this exercise is to say a vowel for as long as possible without changes in intensity.

While software systems may be useful for SLTs, these may not be enough if they are not well designed. It is important to maintain the children's enthusiasm during speech therapy sessions to get better and faster results. Some software systems for speech therapy implement the notion of difficulty levels to challenge the children and keep them eager to practice [4, 3, 11, 2]. The Articulation Station allows the SLT

ACE2016, November 09-12, 2016, Osaka, Japan
 ISBN 978-1-4503-4773-0/16/11...$15.00
DOI: http://dx.doi.org/10.1145/3001773.3001807

to customize the list of words that will be used in the exercise. That way it is possible to increase the level of difficulty by choosing more complex words. On the other hand, Star has predefined levels, where the child should start by teaching syllables to aliens, and as the difficulty level increases, the syllables are replaced by words, and then by phrases. These systems only contain predefined levels or levels controlled by the SLT, so an exercise that may work for one child may not be challenging enough for another child of the same age.

Here we propose a serious game for voice exercises, namely for the sustained vowel exercise, that is controlled by the child's voice in real time. The game extracts relevant speech parameters in order to control the game's main character. This way, by watching the main character's behavior the child has feedback on his performance, and in order to achieve the game's main goal, the child has to try to perform the speech exercise correctly. In addition, the game includes different difficulty levels that can be parameterized by the SLT. This allows the SLT to create more adequate levels for each child's needs.

THE VOICE-CONTROLLED SUSTAINED VOWEL GAME

The sustained vowel exercise is used by voice professionals (such as singers, actors, and journalists) to learn and practice breath control and how to correctly put the voice, and it is used in singing lessons to improve singers' vocal skills. This exercise is also used in speech therapy sessions for children with voice disorders. For instance, it is used to correct hoarse voices. The exercise is also used with children who speak too softly or to loudly. In this case, the SLTs use this exercise to teach the children how to speak in a louder or softer tone of voice, respectively. The sustained vowel exercise with a higher intensity is intended for children who speak very loudly and in a way that may damage their vocal cords. In this case, SLTs can use it with a high intensity to teach these children how to correctly speak in a high tone of voice without being detrimental to the vocal cords.

The exercise consists of saying a vowel with approximately constant intensity for a few seconds. When used to treat voice disorders, the intensity and duration to be practiced depend on the specific voice disorder. For instance, a child who constantly uses a very loud tone of voice and a child who speaks in a whispering mode, should practice the exercise at different intensity levels.

While this is an important exercise for voice disorders, it can be a tedious activity. As a contribution to motivate children during speech therapy sessions, we propose a voice-controlled serious game for the sustained vowel exercise. Our aim is to help SLTs to make more attractive sessions for children through the game. The goal of this game is to have the child enjoy the exercise and thus to motivate him to try to perform better.

The game (created in Unity3D) uses an attractive and colorful scenario and an appealing character, the bird in figure 1. When the game starts, the bird is sitting on the left tree branch. The goal of the game is to make the bird reach the nest on the right branch. To make the bird fly to the nest

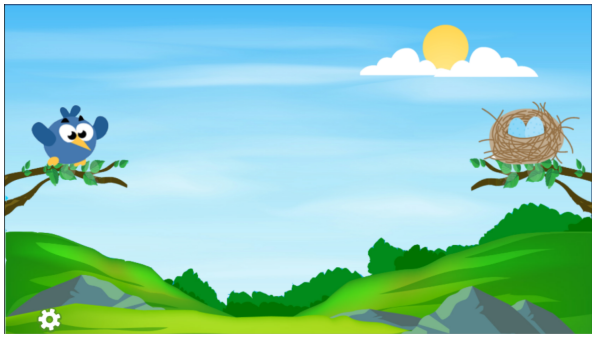


Figure 1: The sustained vowel bird's game.

Figure 2: Game setup.

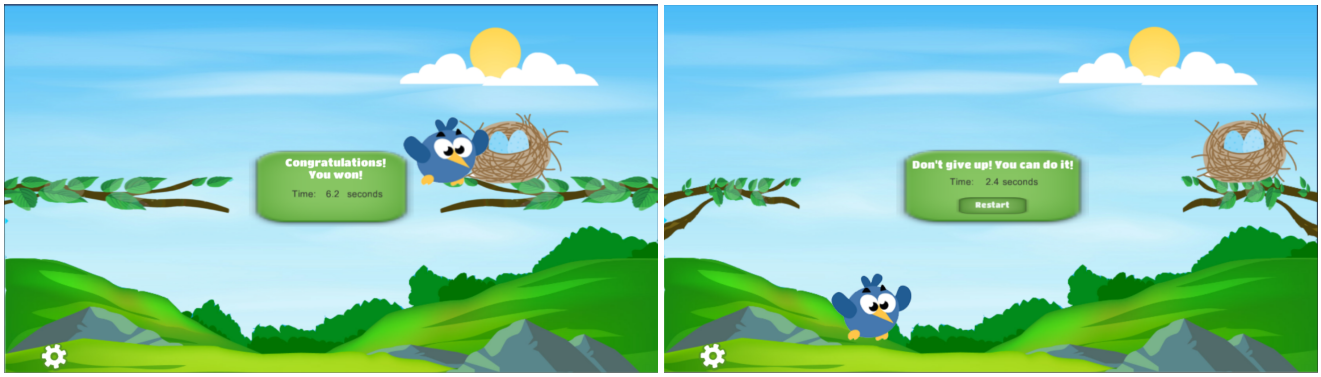
the child has to use her voice. No keyboard, joystick or other input devices are needed. The child has only to use her voice to play the game. As shown in the game setup picture (figure 2), to play the game one needs only a computer and a microphone, which captures the child's voice in real time.

While the child performs the sustained vowel exercise, that is, while the child holds the vowel in an adequate intensity the bird flies towards the right branch. If the child can hold the vowel long enough the bird will reach the nest. When the voice intensity is not adequate or the child stops saying the vowel before the bird reaches the nest, the bird falls.

Voice interaction with the game

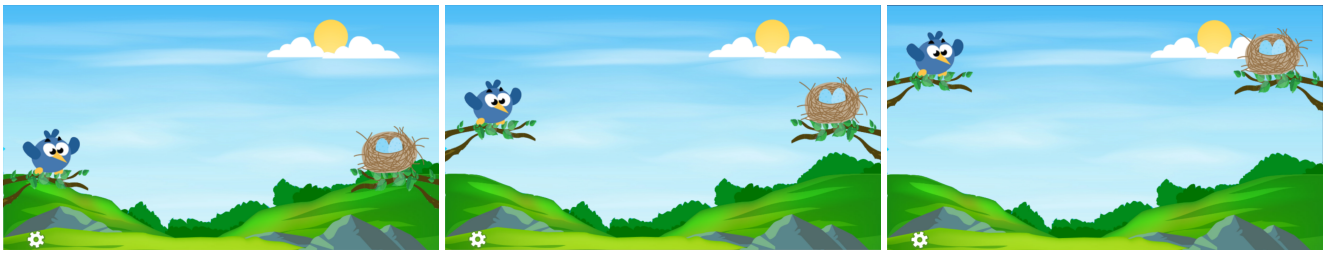
To control the game's main character, we need to extract relevant speech parameters. Since we want the child to sustain the vowel in a specific level of intensity and for a certain time interval, we need to extract the intensity value of the speech production and the maximum phonation time (MPT), where MPT is the amount of time that the child is able to sustain the vowel. To give immediate feedback, indicating whether or not the child is doing the exercise well, the extraction of these parameters must be done in real time.

In this exercise it is very important to control the intensity level, but it is hard to keep it constant. Thus, with the help



(a) (b)

Figure 3: End of game. The green box indicates the phonation time achieved. (a) Bird reaches the nest. (b) Bird falls.



(a) (b) (c)

Figure 4: Intensity levels: (a) low intensity, (b) medium intensity, (c) high intensity.

of an SLT, we have established a minimum intensity level threshold t_m and maximum intensity level threshold t_M to ensure that there is no great variation in intensity during the exercise but at the same time minimum variations are allowed.

During the game, the intensity level is extracted at regular time intervals. If it is between the defined minimum and maximum threshold, the main character keeps moving until it reaches the goal (figure 3.a), otherwise the game stops and the main character can no longer continue, more specifically, the bird falls without reaching the nest (figure 3.b).

Voice intensity levels

Different children have different levels of a specific pathology. For instance, some children talk too softly, while others talk too loudly or even yell instead of speaking. Therefore, besides controlling that the child keeps the vowel intensity approximately constant, it is also important to define what that intensity should be.

With the suggestion of an SLT, and to make the exercise more complete, the game includes three levels of intensity. Thus, the child can practice the sustained vowel exercise at a low intensity, a medium intensity, which corresponds to the normal spoken voice intensity, or at a higher intensity (figure 4). Figure 5 shows the dialog box used to choose the intensity level of the exercise. The SLT uses this box to define the adequate level of intensity of the exercise for each child. The lower slider in the box is used to choose between low, medium and high intensity.

As explained above, the game uses intensity level thresholds (t_m and t_M), to allow small variations in the voice's intensity level. The values of these thresholds depend on the level of intensity chosen for the exercise (in the dialog box of figure 5). Since quiet voice levels, like voice levels at a quiet library, corresponds to 35-40 dB, the low intensity level exercise uses a minimum threshold of 35 dB and a maximum threshold of 40 dB. The medium intensity levels correspond to normal conversation levels, between 40 and 60 dB, which were the values chosen for the thresholds for this case. The higher intensity level used in the game corresponds to values between 80 and 90 dB. All values were recommended by an SLT and a summary is presented in table 1.

Exercise intensity level	t_m	t_M
low	35 dB	40 dB
medium	40 dB	60 dB
high	80 dB	90 dB

Table 1: Values of the minimum and maximum intensity level thresholds.

At the game's current version, the values of the thresholds (table 1) are fixed. However, we are improving the game so that the thresholds can adapt to each child. When the child starts the exercise, we will measure the intensity variation range of his voice and will use it in the definition of t_m and t_M. As

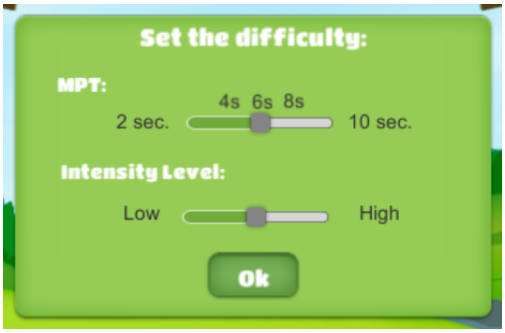


Figure 5: Dialog box to parameterize the expected maximum phonation time and the intensity level.

the child keeps practicing and shows progress and improvements in the exercise, we will slowly change the values of the thresholds to values closer to those presented in table 1. This way, we will avoid that the child gets frustrated in his first trials even if his voice intensity is outside the ranges presented in the table.

Parametrization of the difficulty level

SLTs work with several children of different ages and different SSDs, thus, some children may have more difficulty in performing the exercise than others. Younger children may not be able to sustain the vowel for as long as older children, or some children may have an SSD less severe than others and have less difficulty on performing the exercise. For this reason, it is extremely important to give the SLT the possibility to define the expected phonation time for each child. The proposed game allows the SLT to choose this value through the upper slider in the dialog box shown in figure 5.

The distance between the branches changes according to the expected phonation time chosen. As shown in the figure 6, with an expected phonation time of 10 seconds (which is the MPT estimated for children [6]) the distance between the branches is bigger than with an expected phonation time of 6 seconds. The SLT can choose from five different values for the expected phonation time: 2, 4, 6, 8, or 10 seconds, being that the distance between the branches is proportional to the chosen value.

CONCLUSIONS

Here we propose a serious game for the sustained vowel exercise, which can be used during speech therapy sessions by children with voice disorders. The main goal of the game is to help the SLT on keeping the child motivated to perform the exercise, so as to reach improvements faster than with traditional methods.

The game is controlled by the child's voice in real time, in other words, instead of using the keyboard or other input device to control the main character (the bird) the child uses his voice. This effect is reached by extracting and analyzing some speech parameters in real time: intensity and phonation time. In order to make the bird reach the goal of the game (the nest) the child has to perform the sustained vowels exercise correctly, which is a way of motivating the child to

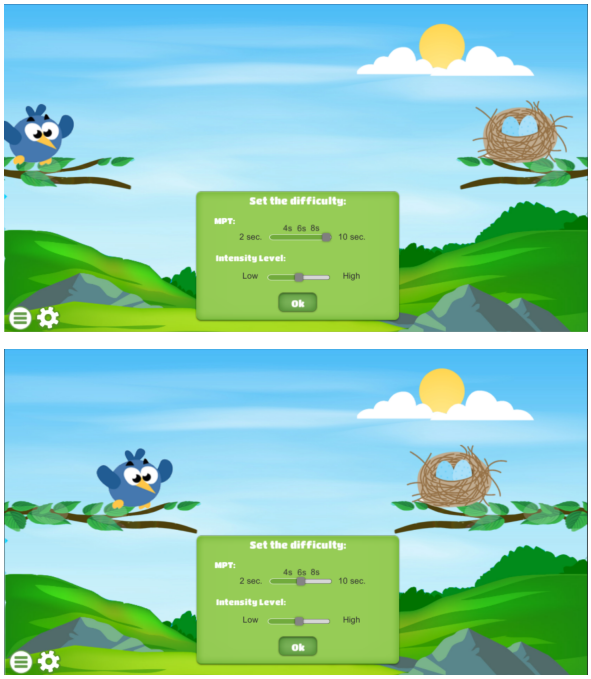


Figure 6: Expected maximum phonation time: (top) 10 seconds, (bottom) 6 seconds.

perform the exercise correctly and even to try to improve. In addition, the bird's behavior gives the child feedback on his performance. This way the child understands if he his performing the exercise correctly and it is easier for the child to understand how to improve his own performance.

Children are different from each other, and different children can have different levels of a specific pathology. Therefore it is desirable that the game can be adjusted to each child needs. Some children can talk too softly while others can talk too loudly. Therefore, different children may have the need to practice the sustained vowel exercise at different intensity levels. As a response, the game includes three levels of intensity: low, medium, and higher intensity levels.

Children of different ages and different SSD, may have different MPTs. Thus it is normal that some children have more difficulty to perform the exercise than others. So we give the SLT the possibility to parameterize the game's difficulty level, by specifying the expected maximum phonation time.

By providing the possibility to adapt the game's difficulty and intensity levels, we give the SLT full control to adjust the game to each individual child. Also, in order to keep the children motivated on practicing and avoid their frustration (when the exercise is too difficult for them) we are introducing ways of adapting the difficulty level automatically, so that it changes in accordance to the child's performance.

During the development of a serious game for speech therapy, it is very important to understand the SLTs' real needs, as they are the ones who will be using the game in the therapy sessions. Having this in mind, there was a constant collaboration

with SLTs during all the phases of the game development. We have received very valuable feedback from them.

In summary, the main novelty of this software application is the gamification of the speech therapy exercises with characters controlled by voice in real time, and parameterization of the difficulty level to each child's needs. As future work, we will offer the option of having the levels of difficulty (expected phonation times) change automatically in accordance to the child's performance. Also, we are working on more scenarios, game goals, and rewards to avoid that the child gets tired of the exercise.

Acknowledgments

This work was supported by the Portuguese Foundation for Science and Technology under projects BioVisualSpeech (CMUP-ERI/TIC/0033/2014) and NOVA-LINCS (PEest/UID/CEC/04516/2013).

Many thanks to Rosário Vieira, for the feedback and help given throughout all phases of the game's development. We thank Tiago Cardoso and his research team for help with technical details.

REFERENCES

1. Abad, A., Pompili, A., Costa, A., and Trancoso, I. Automatic word naming recognition for treatment and assessment of aphasia. In *INTERSPEECH* (2012).
2. Bunnell, H. T., Yarrington, D. M., and Polikoff, J. B. STAR: articulation training for young children. In *ICSLP* (2000).
3. Lan, T., Aryal, S., Ahmed, B., Ballard, K., Gutierrez-Osuna, R., and Texas, A. Flappy Voice: An Interactive Game for Childhood Apraxia of Speech Therapy. In *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-human Interaction in Play*, vol. 2 (2014), 429–430.
4. Little bee speech, 2016. `http://littlebeespeech.com/`.
5. Parnandi, A., Karappa, V., Lan, T., Shahin, M., McKechnie, J., Ballard, K., Ahmed, B., and Gutierrez-Osuna, R. Development of a Remote Therapy Tool for Childhood Apraxia of Speech. *ACM Trans. Access. Comput. 7*, 3 (2015), 10:1—-10:23.
6. Prater, R. J., and Swift, R. *Manual de Terapeutica de la voz*. Masson-Litle, Brown.
7. Rubin, Z., and Kurniawan, S. Speech adventure: Using speech recognition for cleft speech therapy. *Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments - PETRA '13* (2013).
8. Soto-Faraco, S., Calabresi, M., Navarra, J., Werker, J., and Lewkowicz, D. J. The development of audiovisual speech perception. *Multisensory development. Oxford University Press, Oxford* (2012), 207–228.
9. Speech-trainer.com, 2016. `http://speech-trainer.com/children-speech-`.
10. Tan, C. T., Johnston, A., Ballard, K., Ferguson, S., and Perera-Schulz, D. sPeAK-MAN. *Proceedings of The 9th Australasian Conference on Interactive Entertainment Matters of Life and Death - IE '13* (2013), 1–4.
11. Tavares, J., Lopes, J., Cunha, M., and Saldanha, R. Falarabrincar.wordpress.com, 2016. `https://falarabrincar.wordpress.com`.