Integrating Computer Games in Speech Therapy for Children who Stutter

Daniil Umanski
Leiden University Centre for Linguistics
Leiden Institute for Brain and Cognition
Leiden, The Netherland
+31624726658
rabotnik@gmail.com

Walter Kosters, Fons Verbeek
Leiden Institute of Advanced Computer Science
Leiden, The Netherlands
kosters@liacs.nl
fverbeek@liacs.nl

Niels O. Schiller
Leiden Institute for Brain and Cognition
Leiden, The Netherlands
n.o.schiller@let.leidenuniv.nl

ABSTRACT
In this paper we describe our work with the development of a novel computer game supporting speech therapy for children who stutter. We discuss the motivation for our work, the theoretical background, and outline the plans and strategies for the development process. Finally, we describe a preliminary study carried out to evaluate the potential of integrating computer games in speech therapy for children.

1. INTRODUCTION
Stuttering is a speech disorder in which the flow of speech is disrupted by involuntary repetitions, prolongations of sounds, and silent blocks. From a motor point of view, stuttering can be described as a disorder in the timing and coordination of the subsystems involved in speech production, namely respiration, phonation, and articulation [7]. One of the most common forms of stuttering treatment for children is speech motor training. In a speech clinic, a child is typically asked to repeat a large amount of syllable strings, such as "puh-tuh-kuh". By systematically manipulating the complexity of syllables, speech motor training aims at establishing correct speech motor programs at the syllable level. After training, a child uses those motor programs during spontaneous speech. While training, the speech therapist provides verbal feedback on the child's production, and directs to the next exercise. This form of practice presents several inherent difficulties:

» A child is expected to remain motivated to perform large amount of practice, while pursuing tasks which are rather abstract by nature.

» The feedback is provided to the child only after he/she has produced the syllable set. This does not allow the child to monitor specific articulatory movements in real-time and learn to correct.

» Training is dependent on the presence of a speech therapist, which is expensive. A child has no possibility to practice the learned skills at home at a preferred moment of time.

2. APPROACH
One potential tool to improve the current training situation is a computer-based system. Considering the remarkable fascination of children with computer games, it seems well-reasoned to house speech motor training within a computer-game environment. Educational computer games have been widely recognized as having great potential to motivate children to perform tasks which are not intrinsically motivating. This is especially relevant for children who need to practice certain skills where a disability is diagnosed. The usage of new media enables developers to engage children in learning not only by boosting motivation, but also by creating conceptually novel tasks, which may go beyond the paper-based training methods. The major challenge of training-system developers is to seemingly incorporate relevant educational or therapeutic elements into the "world" of the game.

If such elements trigger learning mechanisms of particular skill while taking advantage of the increased attention and motivation levels of the player, a potentially highly effective training tool can be offered. According to Hubbard [5], attractiveness should be a primal concern when designing educational software. He suggests that while playing computer games, the immersion effect is created, whereby the environment into which the players are submerged progressively increases their levels of attention and concentration on the goal, which stimulates the learning process. The power of computer games is in providing immediate and accurate feedback in a systematic way. As opposed to a delayed form of feedback, a child perceives the correlation of his/hers performance to the presented goal in real-time, and gains the ability to monitor and adjust his/hers behaviour in a very precise manner. This immediate correlation is central to the speech training environment we are developing.

The issue of control plays a major role in the design of interactive programs for children. The common approach is to allow children to choose between different formats of exercise and different levels of difficulty. In general, research indicates that the key points for creating effective training computer games are provision of feedback, control, curiosity and a feeling of competence [8]. To create the feeling of challenge, goals must be clearly presented, and be not too easy nor too difficult. When children try to reach a specific goal, they must be provided with an intuitive feedback. Curiosity is recognized as a powerful source of intrinsic motivation [8]. A structured build-up progress through difficulty levels will increase motivation to go further in the training. The idea is to channel existing computer-usage habits of children towards a useful and learning-aimed interaction.

With those ideas in mind, together with speech therapists we have decided to develop a new training tool for children who stutter. The training tool is implemented as a computer game, in which the child is encouraged to achieve specific goals by using his/hers voice to play the game. The game provides real-time visual feedback on the quality of speech productions and motivates the child to improve speech motor skills in order to progress in the game. The feedback we plan to provide is related to the bio-feedback paradigm, where a certain physiological signal is visualized in real-time to allow a person to gain some amount of control over it [2]. In our case the bio-signals are the voice parameters relevant to the specific speech motor training methodology. The structure of exercises used to train speech motor skills is reflected in the game, which becomes a dynamic extension of the treatment program. The therapy game is by no means assumed to replace therapy sessions with a clinician, but rather to complement existing therapy, and provide an attractive possibility to practise speech motor skills at home.

3. SYSTEM DEVELOPMENT

3.1 Training methodology

An important aspect of speech motor training methodology is the choice of verbal stimuli to be practised. Researchers distinguish between stimuli of real words, non-sense words which have a legal word structure, and non-sense syllable strings which have an unnatural prosodic structure, since they are equally stressed. It is agreed that the non-sense syllables stimulus yields a measure most related to speech motor abilities, since it is least affected by linguistic factors which influence real-word and "legal" non-sense word repetitions. Since the syllable string is completely unfamiliar, most likely never heard or produced before, its production cannot rely on prior linguistic knowledge, and a child has no stored motor programs to rely on. Therefore, this kind of practice requires an assembly of new motor programs, which allows researchers to focus on measuring and training the speech motor skills per se.

We plan to take advantage of psycholinguistic studies about the "mental syllabary" [1] to contribute to the methodology underlying speech motor training. This syllabary is assumed to contain pre-compiled articulatory motor programs for the most frequently used syllables in order to facilitate their rapid production, such that speakers do not have to assemble these syllabic motor programs from scratch each time they produce them. To support those assumptions, Schiller et al., [6] have found that for English, Dutch and German, approximately 80% of speech in those languages utilizes only 500 different syllables, which is only 5% of the entire syllable inventory. This suggests that speakers produce the same syllables over and over again. Therefore, training children who stutter with syllables from different "frequency bins" in a structured program should produce different effects on speech motor skills. More precisely, training syllables that occur often in the language of the speaker should produce a greater advantage in daily language use for the speaker than syllables that rarely occur in the language.

3.2 Design strategy

Our starting points in developing the new training system are the following key questions regarding system usability:

» How to make the system adaptive to different users?

» Which age groups can benefit from the therapy game ?

» Which form of feedback best stimulates children to practice?

In the process of design and implementation, the aim is to draw from existing research in developmental psychology dealing with learning and child-computer interaction. This knowledge, together with clinical experience is used to realize the new training system. The departure point is interviewing speech therapists in order to elicit requirements for the new system. Once the concepts for the system are formulated, the research-through-design approach is taken. The process involves several cycles of prototyping and testing for usability with children (not only children who stutter). Our guidelines for designing a sustainable therapy tool are:

» Adaptive mechanisms to support individual, child-specific training procedures. One possibility is to allow the clinician to set child-specific goals and training sets. Another option is to program the system to adapt dynamically to the performance of the user by adjusting targets on the fly.

» Key partners in the process of game development are the children themselves. Our experience so far echoes the enormous value of children's contribution to the design and evaluation of interactive computer games (see also [4]).

» Sensible speech analysis and visualization. The game should provide visual feedback on relevant speech parameters, such as intensity and pitch contour, voiced/voiceless consonant detection, smooth coarticulation, correct rate and rhythm.

3.3 Speech signal processing

Following the need for a more intelligent analysis of the incoming speech signal, we are now investigating the possibilities of integrating a machine learning component based on speech recognition (SR) technology. The idea to utilise a recogniser is related to the desire of providing feedback on the quality of specific phonemes or syllables which will be part of the training program. Since we know in advance which syllable sequences are expected, the SR component could be able to segment the incoming speech signal according to the expected syllable sequence, which means to align the two sequences with markers on segment borders. Next, the segments would have to be evaluated for the relevant parameters (precision of articulation, substitution of phonemes, smoothness of coarticulation etc.) according to the acoustic models corresponding to the expected segments. This in fact is an elaborate pattern recognition process which can potentially yield quality scores (confidence/distance measures) for our exercises. However, in order to perform such elaborate analysis, the system must be first trained on a rather large database of speech samples, representative of the expected input tokens. Furthermore, next to each syllable sequence production, such a database must contain a transcription of the relevant error types, as well as clinician's judgements of the sequence. The compilation of such an elaborate database is logistically rather non-trivial, demanding many expert working hours. This disadvantage is common to supervised machine learning techniques. In general, those methods need a fully annotated training database to learn to discriminate incoming signals. Alternatively, we can think of looking on the fly at the acoustic features of the incoming signal, having a set of acoustic filters (or discriminators) to decide on the overall quality. The disadvantage of such an approach is that the system is not learning from experience.

A reasonable solution would be to conceive an unsupervised learning routine, in which a database of speech samples (not fully annotated, but with overall quality scores) is passed through a set of acoustic-based discriminators (a sequence of acoustic analysis functions). The resulting data structure is a n-dimensional space, where n = number of discriminators. A classification function is then derived which separates the data-space into areas of quality scores. Further on, in the real-time situation, the same set of acoustic discriminators is applied to the incoming speech signal, and the classification function is used to obtain a quality score. The advantage of such an approach is that machine learning is applied to incorporate previous knowledge of utterance judgement (by clinicians) to the evaluation of productions in real time. The database needed to train such a system needs not to be fully annotated, which makes its compilation much more feasible.

4. SYSTEM VERIFICATION

4.1 Usability studies

After each design iteration, the developed prototype will be tested with children. The usability tests will aim to answer the posed research questions — revealing which age groups are most suited to use the system, and how motivated children are to interact with the current prototype. Following guidelines of testing usability in child-computer interaction [3], we will conduct qualitative studies by means of:

» Questionnaires — items on motivation, attitude, appreciation.

» Behavioural observations — video recordings and independent judgements about engagement.

» Free choice — frequency of selection and the time spent with the speech-training system, or with two other games.

4.2 Treatment efficacy studies

Once the developed training system is "mature", we will conduct treatment efficacy studies with children who stutter. Together with speech therapists, we will measure the effect of training with the new system on stuttering severity of children. In the process of testing the effectiveness of the developed system, questions will be raised concerning the effects of training on fluency:

» How does training affect spontaneous speech of children who stutter in long the term?

» How does an individual child's stuttering history (time since onset of stuttering, severity of stuttering) affect the treatment outcomes for that child?

» Which syllable sets are most effective as training material?

The treatment efficacy studies will utilize a group comparison design. One group of children who stutter will undergo speech motor training with the computer-game, and a control group will receive standard speech motor therapy. A total number of 30 subjects will be recruited, aged 6, 7, 8, and 9 years. The training will last for 15 weeks. Measurements will take place pre-treatment, 5 and 10 weeks into treatment, and post-treatment. Dependent variables are stuttering frequency, measured as % SS = percentage of stuttered syllables (see Figure 1), and stuttering severity rating (a continuous scale from 0 to 7). Audio and video

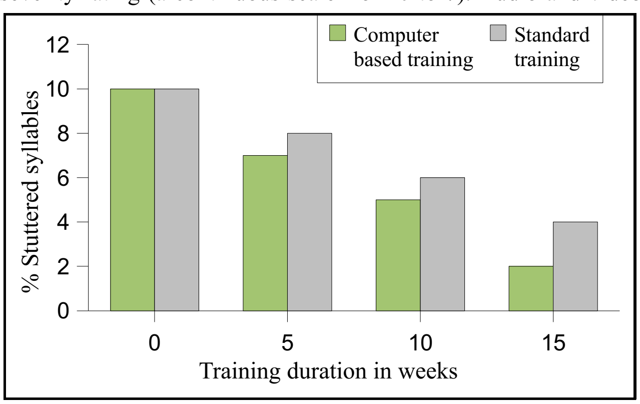

Figure 1: Exemplary fluency measurements: percentage of stuttered syllables over time.

samples of children's productions will be collected throughout various speech tasks. Those tasks vary in complexity from picture naming tasks, reading tasks, and finally spontaneous speech. Samples will be analysed with the help of a research assistant, and stuttering frequencies will be measured independently by two researchers in order to provide high reliability. The stuttering severity rate will be perceptually determined by a speech therapist and the parents of a child.

4.3 Psycholinguistic studies

A psycholinguistic study will measure how treatment outcomes are affected by manipulating different syllable sets for training. The aim of the psycholinguistic study is to reveal which choice of syllable sets will provide the most efficient training material. In the empirical study, we will parametrically vary the frequency of occurrence of the to-be-trained syllables while controlling for potentially confounding factors such as segment frequency and articulatory difficulty arising from the combination of certain segments (co-articulation). Productions of various target syllables will be evaluated using quantitative criteria such as voice onset latencies and speech errors by professional speech therapists before and after the training.

5. PRELIMINARY STUDY

5.1 A prototype game

So far, we have conducted an initial study on the potential of using a therapy game to supplement existing forms of speech motor training. In that study, we have built and experimented with a straightforward prototype of a speech training game [9]. In a training session, a child controls the motion of an antagonist figure using his/her voice, with the goal of hitting targets on the screen. Those targets represent the syllables which the child has to produce, so that targets in the game are hit when the child

Figure 2: A screen shot from a prototype game.

produces the correct speech pattern in terms of rhythm and continuous phonation (see Figure 2).

A rather basic analysis is used by applying a low-pass filter to the intensity signal so that a smooth loudness pattern is obtained and can be visualised. This object returns the loudness level of the incoming signal based on the psycho-acoustic approach to human sound perception. This is found to be the most practical and robust approach in application-driven work. The loudness is the subjective judgement of the intensity of sound. First, the Short Time Fourier Transform is applied to a window of size 512, and the spectrogram is obtained, containing the energy levels of 512 frequency bands. Subsequently, each energy value E_k is summed up and the average constitutes the loudness at time t (see formula below).

$$L_{dB}(t) = \frac{\sum_{k=1}^{N} E_k(t)}{N}$$

where E_k is the amplitude of frequency band k of total N in the spectrogram (N = 512). This approach allows us to catch even subtle pauses in phonation and by this discriminating relatively continuous and abrupt voicing.

Short gaps in the signal intensity are detected, assuming to reflect breaks in phonation, enabling the program to trigger appropriate visual cues. In the current version, the wooden floor underneath the game hero would break at the point where the break in phonation was detected. The game prototype does not include any mechanism to distinguish and recognise specific phonemes or syllables. After each exercise, the score is presented and the child proceeds to the next level. The clinician can set child-specific difficulty level, rate of the exercise, and switch between different modes of visualisation (see Figure 3).

5.2 Clinical tests

We have tested the prototype game in a speech clinic with children who stutter. We have conducted two experiment sessions with children at the Rijnland Stuttering Clinic in Oegstgeest, The Netherlands. In each session we have tested with 8 children, one at a time. There were 2 girls and 14 boys. We have invited both children who stutter and those who do not. In total, 4 out of 16 children were non-stutterers. We wanted to check if those children relate differently to the game than those who stutter. The ages of the children in the first session were: 9, 11, 12, 9, 11, 4, 7, 8. Children in the second session were: 6, 10, 8, 6, 10, 5, 9, 9.
The main question we have posed is whether children relate easily to the kind of interaction the game offers. The goals of those experiments were as follows:

» Observe the reaction of children to the prototype of our game.
» Observe whether children show motivation to succeed.
» Collect explicit impressions of children about their experience.
» Look at the way clinicians refer to the game as a therapy tool.
» Observe the reaction of parents to the training process.
» Analyse the strong and weak points of the development so far.

5.3 Results

The overall reaction of children to the game prototype was positive. This impression was built from both observing the children at play time and asking them explicit questions afterwards. It was clear to see that in almost all cases, children seemed involved and focused on the exercise. Another clear point was that the children immediately understood the principle and rules of the game without explicit explanations. In order to evaluate how intuitive the interaction with the game is, we have simply let a child play the game without instructions and guidance. After the first few trials with their voice, all children picked up the principle and started aiming to hit the target objects while pronouncing the syllable sets. We have recorded no case of a child asking how does the game work or being puzzled by the reaction of the moving figure to his/her voice or by the way the target objects react to the figure hitting them. Based on a clear pattern of immediate relation to the game's mechanism, we can conclude that children found the interaction principle simple and intuitive. Observing the determination expressed by most children during the syllable production as well as the joyful reactions upon completing a screen with optimal performance, we can assume they experience a rather high motivation to succeed. A clear contribution to the motivation factor were the scoring screens of the game during the second test session. The children could perceive in a clear way how well they had performed and reacted strongly either with satisfaction or dissatisfaction, and in both cases with an intention to perform better. This observation goes along well with our prediction that a concrete and simple goal would create motivation to succeed in the exercises.
A very important factor on the child's approach to the game proved to be the presence of the clinician. It was clear that children were more comfortable with her being present. Her role proved to be highly efficient in guiding a child to interpret the visual feedback of the game and correct the syllable production based on that. For example, she would ask the child “Why the running figure did not reach the target?”, and the child would respond by recognising the relation of his/her voice volume to matching the target. By guiding the child, the clinician is able to refer to specific ways for improving certain voice quality in terms of the game's symbols, to which the child easily relates. For example, instead of saying "Your volume is lower at the last part of the syllable set", the clinician would say "Try to hit the last three cows as well now". Thus, the role of the clinician during the training is important in few aspects, at least in the first stage, when a child gets accustomed to the game. First, making the child comfortable with the training situation. This we have especially observed when the child first made contact with the game. There were a few cases during the second test session, when children refused to approach the game in the first place without the clinician being present. Secondly, at providing guiding cues to the child during the exercises through the symbolic language of the game, thus referring to concrete visual objects and tasks.

Figure 3: An interface screen for the clinician.

6. CONCLUSIONS & FUTURE WORK

The results of our preliminary clinical tests and usability evaluations convincingly suggest that computer games are a powerful tool for motivating children to practice speech motor skills. The experiments showed high attention and concentration levels of children who practised, as well as short term improvement in performance in terms of the game scores. The visual environments used in the prototype game proved to be easy for children to relate to, however more variety is needed to sustain curiosity.
Following our current results and experience, we have sketched the main directions for future work, with the idea to proceed with building a widely usable training tool. Psycholinguistic studies should expand our understanding about the relative effectiveness of various syllable types in speech motor training. Besides providing a firm theoretical framework for the treatment program, such findings will indicate the best choice of training material for the program. Furthermore, we plan to investigate which speech processing techniques can best provide evaluation on the quality of production in a real-time situation. We aim to strike a balance between utilising machine learning approaches while not relying on the existence of a very specific speech corpus.
Finally, a treatment efficacy study should provide quantitative support for the effectiveness of our training system in real-life treatment situations. Once developed and verified, the system will offer clinicians and children a new evidence-based tool to practice speech motor skills in an attractive and playful way.

ACKNOWLEDGEMENTS

This research is funded by a Mosaic grant from the Netherlands Organisation for Scientific Research (NWO).

Our kind gratitude goes to the speech therapists of the VSN (The Dutch Association of Stuttering Clinics) for guiding our development, and to all the children who are involved in the project.

REFERENCES

[1] Cholin, J., Levelt, W. J. M., & Schiller, N. O. “Effects of syllable frequency in speech production”. *Cognition*, 99, 205-235 (2006).

[2] Darby, Mc. , Condron, G., Hughes, G., AugenBlick, N. “Affective feedback” Submission chapter for a book entitled: *Enabling Technologies*, Elsevier Science Ltd. (2001).

[3] Donker, A., & Reitsma, P. ” Usability testing with young children”. *Proceedings of IDC 2004.* pp 43-48. New York: ACM Press (2004).

[4] Druin, A. “The role of children in the design of new technology.” *Behaviour & Information Technology*, 21(1), 1-25 (2000).

[5] Hubbard. “Evaluating computer games for language learning”. *Simulation and Gaming* 22, 220–223 (1991).

[6] Schiller, N. O., Meyer, A. S., Baayen, R. H., & Levelt, W. J. M. “A comparison of lexeme and speech syllables in Dutch.” *Journal of Quantitative Linguistics*, 3, 8–28 (1996).

[7] Kent, R.D. “Research on speech motor control and its disorders, a review and prospective.” *Journal of Communication Disorders* 33, 391-428 (2000).

[8] Ryan, R.M. & Deci, E.L. “Intrinsic and extrinsic motivations: classic definitions and new directions.” *Contemporary Educational Psychology*, 25, 54-67 (2000a).

[9] Umanski, D. “Computer game development for speech therapy support.” *MSc thesis,* Leiden University (2006).