On the Development of an ASR-based Multimedia Game for Speech Therapy: Preliminary Results

Mario Ganzeboom
CLS/CLST, Radboud
University Nijmegen
Erasmuslaan 1
Nijmegen, The Netherlands
m.ganzeboom@let.ru.nl

Emre Yılmaz
CLS/CLST, Radboud
University Nijmegen
Erasmuslaan 1
Nijmegen, The Netherlands
e.yilmaz@let.ru.nl

Catia Cucchiarini
CLS/CLST, Radboud
University Nijmegen
Erasmuslaan 1
Nijmegen, The Netherlands
c.cucchiarini@let.ru.nl

Helmer Strik
CLS/CLST, Radboud
University Nijmegen
Erasmuslaan 1
Nijmegen, The Netherlands
w.strik@let.ru.nl

ABSTRACT

A potential consequence of the ageing population is an increased incidence of neurological diseases that cause communicative disorders. In turn, this may lead to an increasing demand of intensive and costly speech therapy. To alleviate this problem, multimedia applications in the area of telerehabilitation and web-based speech training have been developed to support speech therapy. However, due to the repetitive nature of some exercises, therapy is not always perceived as particularly motivating. This paper reports on research aimed at developing a multimedia game that incorporates Automatic Speech Recognition (ASR) technology to provide patients autonomous and motivating practice without the intervention of a speech therapist. Currently, the game includes visual feedback on two dimensions of dysarthric speech that often deviate from healthy speech. To explore the possibility of integrating feedback on dysarthric speech by using ASR technology, initial experiments were conducted on available speech databases. The results show that employing ASR is becoming feasible thanks to recent developments in acoustic modelling.

Keywords

Multimedia game; Speech therapy; e-Health; Speech rehabilitation

1. INTRODUCTION

One of the potential consequences of an increasingly ageing population worldwide is a growing incidence of neurological disorders such as Parkinson's Disease (PD), Cerebral Vascular Accident (CVA or stroke) and Traumatic Brain Injury (TBI). These diseases are known to cause communicative disorders. One of them is dysarthria, a motor speech disorder that affects speech intelligibility and causes communication problems [10].

Previous research has shown that intensive speech therapy is beneficial towards improving intelligibility [28]. However, with the growing number of dysarthric patients such therapy may be difficult to provide in the future. A compounding problem is that some speech exercises (i.e. 'speech drills') remain repetitive in nature and patients do not find them particularly motivating.

In our project, research is being conducted to explore the role of multimedia games in increasing patients' motivation to practice intensively. We focus our efforts on developing a multimedia game for speech therapy for Dutch elderly individuals with dysarthria (e.g. 55 or above), because these constitute the majority of the patients' group. The game currently provides feedback on voice pitch and loudness, two speech dimensions that are often affected by dysarthria. We are also exploring the possibilities of using Automatic Speech Recognition (ASR) technology to provide automatic feedback on pronunciation, as was done in previous e-Learning and e-Health projects [31].

In this paper we report preliminary results of our research on developing a game that incorporates ASR technology to provide practice and feedback to Dutch dysarthric patients. After discussing background research on multimedia applications and games for speech therapy in Section 2, we organize the paper in two parts. Section 3 describes the process of game development and the game that was eventually developed. Section 4 reports on the experimental setup and results of our speech recognition experiments. Discussion and conclusions are presented in the final section.

2. RELATED WORK

In the past, multimedia applications have been developed to support speech therapy. The term multimedia is often associated with computer-based applications in which different media like text, audio, still and motion pictures are

MMHealth'16, October 16 2016, Amsterdam, Netherlands

ISBN 978-1-4503-4518-7/16/10...$15.00

DOI: http://dx.doi.org/10.1145/2985766.2985771

used, but also other information from input media like keyboard, touchscreen and microphone. Initially, multimedia applications in speech therapy analysed prerecorded speech samples and provided visualisations of speech dimensions (e.g. pitch, loudness, speech rate, etc.) to be interpreted by experts [7]. Later research describes applications that included interactive visualisations designed for the non-expert individuals who received speech training [17, 5, 35]. Teleconferencing multimedia applications provided the additional advantage that individuals receiving speech therapy could stay at home, reducing the number of journeys to a rehabilitation centre [21, 33, 6]. Web-based multimedia applications for speech therapy also provide this advantage and potentially enable intensified speech therapy in the home environment, reducing dependence on the availability of speech therapists [23, 14, 3]. Recently, research has also started to explore the possibilities of mobile applications to assist patients in training at home [34].

Many of the early applications showed feedback to clients in the form of technical diagrams or vocal tract schemes that clients found unappealing or difficult to interpret without considerable instruction [12, 2]. Several researchers reported on their attempts to identify visualisations that were found appealing and easy to understand in providing feedback on speech dimensions [11, 15, 25].

Multimedia applications with interactive games have been applied in the past to make speech therapy more fun and playful. For example, outside of academic research, the software packages Dr. Speech[1] and VoxGames[2] contain games for children with varying speech disorders. Similar sets of small games were described in academic research as well [29, 1]. These games were intended to assist the work in speech therapy focusing on the use of voice activity, intensity, breathing, tone and vocalization of children with hearing impairments or developmental disorders. Speech analysis algorithms were incorporated to provide children real-time feedback on these dimensions by controlling game characters and objects with their voice. Several researchers also developed multimedia games incorporating Automatic Speech Recognition (ASR) to provide feedback on pronunciation exercises [4, 30, 32].

The previously described games were all directed at children. Recently, Krausse et al. [19] reported on a study that incorporated a multimedia game to provide challenging speech therapy to elderly with dysarthria. The game aimed to improve their reduced voice intensity by challenging them to break glasses and vases by producing sufficiently loud and long /a/-phonemes. The patients appeared to be engaged and enthusiastic about continuing playing. Speech analysis algorithms were used to extract voice intensity from the speech signal.

3. PROTOTYPE MULTIMEDIA GAME

3.1 Game development

3.1.1 Initial concepts

In the process of developing the game, interviews and tests with potential users were conducted at different stages. First, different game concepts were tested with three dysarthric patients and with an advisory group consisting of two (other) dysarthric patients and a game expert. The game concepts tested varied along different dimensions: a) whether they put emphasis on the therapy aspect or favoured the game element and b) whether they were single-player or multiplayer. The interviews and tests revealed that the subjects preferred game concepts with emphasis on gaming, without specific attention for the therapeutic aspects. In addition, a multiplayer game concept was considered more appealing. One of our concerns about a multiplayer game was that patients would not be able to practice independently. However, this was not viewed as a problem, while the social and collaborative aspects of a multiplayer game were seen as an important incentive to play the game more often, with consequent advantages for the therapy. Initial multiplayer game concepts showed a possible imbalance in the amount of time the players would speak, with one of them speaking very often and the other significantly less. Since such an imbalance is a potential obstacle to effective speech therapy, we opted for a cooperative playing style in which both players have to help each other. This showed a better balance in later playtests.

3.1.2 User specific developments

Tests of more advanced game concepts with dysarthric patients indicated important elements that had to be taken into account with respect to this group of older people with very limited experience with gaming. First, the need to gradually introduce new game elements and increased levels of difficulties. For example, introducing only one new element every three or four levels instead of one every level. Second, throughout all of our tests we observed that players did not always (fully) read or follow up on textual instructions shown on the screen. In the interviews users told us that reading multiple lines of text requires considerable effort and they had difficulties remembering them.

Third, integrating indirect feedback on voice pitch and loudness that was part of the gameplay appeared difficult for our users to understand and process. Users often got confused because the feedback on these two dimensions occurred at the same time and continuously changed the main game view. A more direct approach in a later concept that did not involve the main view, was found easier to understand and process.

Prototypes of the game were adapted and improved based on the above observations. This continuous process led to the current version of the game which will be presented below.

3.2 Game Concept

The multimedia game that was eventualy developed in our project can be played by Dutch-speaking dysarthric individuals with their relatives and friends. The game is called 'Schatzoekers' (i.e. 'Treasure hunters') and the goal is to navigate a map and find the treasure. It is a cooperative game in which two players try to reach the goal by helping each other in finding the treasure chest and the key to open it. Players have to talk to each other and give directions via a voice chat connection. One player, the 'digger', is only allowed to walk on land in search for the treasure chest. The other player, the 'diver', is limited to stay in the water in search for the key. Figure 1 visualises the instructions shown

[1]See http://www.drspeech.com/SpeechTherapy5.html, last retrieved on April 25th, 2016.

[2]See http://www.ctsinf.com/english/#voxGames.html, last accessed on April 25th, 2016.

before starting a level in the game. The coloured X´s shown in that Figure, mark the locations of the treausure chest and key on the map shown after the players started the level.

Figure 1: Pre-level instructions showing for the 'digger'. The coloured X´s depict the locations of the treasure chest and key on the map.

However, only the 'digger' can see the location of the key (where the 'diver' should go) and only the 'diver' can see the location of the treasure chest (where the 'digger' should go). The players thus have to explain to each other where to go. This way, they are encouraged to keep speaking to each other by describing their location on the map and providing directions how the other should proceed. Figure 2 shows a screenshot of a level in the game. Every level is a different map and levels get more difficult as the players progress. As some of our players may also be affected in their cognitive skills due to their neurological disorder, it is challenging to find a balance in difficulty. That is why the first levels were designed to introduce the game and slowly increase the difficulty by larger maps and removing the ability to see the other player on the map as is shown in Figure 2. The bottom-right button in Figure 2 provides access to an overview of the full map in the current level. This can initially be used by players for orientation and to discover where the other player can find the treasure chest or key. In later levels, the overall difficulty is also increased by initially removing the locations of the treasure chest and key from the overview. The icons providing orientation information (e.g. windmill, flat building, factory, etc.) are removed next and lastly, players must do without the overview.

The therapeutic goals of the game are to motivate dysarthric individuals in using continuous speech, and to speak up and maintain predefined levels of voice pitch and loudness. Players talk to each other using headsets and the game provides feedback on pitch and loudness by analysing their speech signals. This is indicated by the horizontal green bar shown at the top in Figure 2 which provides real-time feedback while speaking and shows a green, orange or red color when the loudness is within, near or below the threshold, respectively.

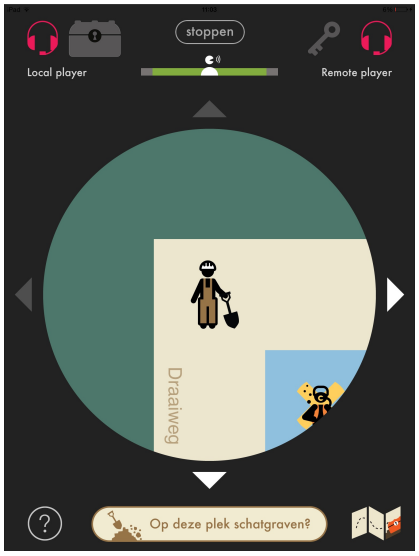

Figure 2: An in-game screenshot displaying the game from the perspective of the player 'digger'. In the partially visible square at the bottom-right of the circle it can be observed that the 'diver' already reached the correct location of the key.

When the pitch is too high a notification slides down from underneath the bar instructing the player to 'speak loud and low'.

To conclude this subsection we highlight the following aspects of the game. Using the concept of having to provide directions to each other, we implemented our first goal to let players use continuous speech. By providing visual feedback based on the automatic analysis of players' speech, we implemented our second goal to encourage players to maintain predefined levels of voice pitch and loudness.

3.3 Platform Architecture

Previous multimedia games were mostly developed for desktop platforms. However, the current widespread availabilty of mobile platforms offers a number of advantages. As a matter of fact, these are more easily accessible and stimulate users to practice more often, which is in line with our goal of providing intensified speech therapy. Additionally, their style of interaction, mostly via touchscreen, is perceived as easier than using keyboard and mouse. For these reasons the present game was developed for the Apple iPad tablet platform.

As one of the goals of the game is to encourage players to maintain predefined levels of voice pitch and loudness, we also developed a software architecture that enables us to change these levels from a distance, if required. In addition, a logging system was developed for recording all user-game interactions as well as the speech produced by the users while playing the game. The speech recordings provide a valuable resource for further research on dysarthric speech and therapy effectiveness and for developing technology-based applications for speech therapy. Additional logging of user actions potentially provides an understanding of how the game is used. In current efforts to explore the possibilities of using ASR technology for pronunciation feedback, we also incorporated a speech recogniser. The software architecture

currently runs on one physical machine at a cloud server provider and spawns a separate session for every player including a dedicated recogniser instance. This allows us to easily offload sessions to additional machines if necessary.

4. RECOGNITION EXPERIMENTS

4.1 Experimental Setup

4.1.1 Speech Data

One of the aims of our research is to use the game to provide more informative feedback on dysarthtic speech by using ASR technology for speech analysis. Ideally, the ASR component to be incorporated in the game should be developed and tested on the speech that is produced in the game. However, this would introduce a sequentiality in the development process that is not feasible in a project with limited duration. For this reason, we decided to perform ASR recognition experiments on speech databases that were already available and were considered to be relevant to our purpose. One of these is the recently created Dutch dysarthric speech database [36], which contains 6 hours and 16 minutes of dysarthric speech material from 16 speakers. The speech segments with pronunciation errors were excluded from the test set to maintain integrity of the results on ASR performance evaluation. In addtion, the segments including a single word and pseudoword were also excluded, since the sentence reading tasks are more relevant for our multimedia game. The duration of the final test data is 4 hours and 47 minutes. Based on the meta-information, the age of the speakers is in the range of 34 to 75 years with a median of 66.5 years. The database contains dysarthric speech from ten PD patients, four CVA patients, one TBI patient and one patient with a birth defect. The level of dysarthria varies from mild to moderate.

Aiming to cope better with the speech deviations caused by dysarthria, we trained deep neural network (DNN)-based acoustic models altering the number of hidden layers to explore their impact on the recognition performance. For training the acoustic models, the CGN corpus [24] was used, which contains a representative collection of contemporary standard Dutch as spoken by adults in the Netherlands. The components with read speech, spontaneous conversations, interviews and discussions were used for training the acoustic models in the present experiments. The duration of the normal Dutch speech data used for training is 255 hours and 11 minutes. Taking the median age of 66.5 years into account, elderly speech data from the JASMIN corpus [8], was added to normal speech in the training phase. The duration of that speech is 10 hours and 10 minutes.

4.1.2 Implementation details

The recognition experiments were performed using the Kaldi ASR toolkit [27]. The standard training recipe provided for multiple databases was applied to train a conventional context-dependent GMM-HMM on FMLLR-adapted features. A standard feature extraction scheme was used by applying Hamming windowing with a frame length of 25 ms and frame shift of 10 ms. An SGMM-based [26] acoustic model was also trained with a universal background model having 800 Gaussians and substate phone-specific vector size of 40. This system was used to obtain the state alignments required for training the DNN-based recognizer.

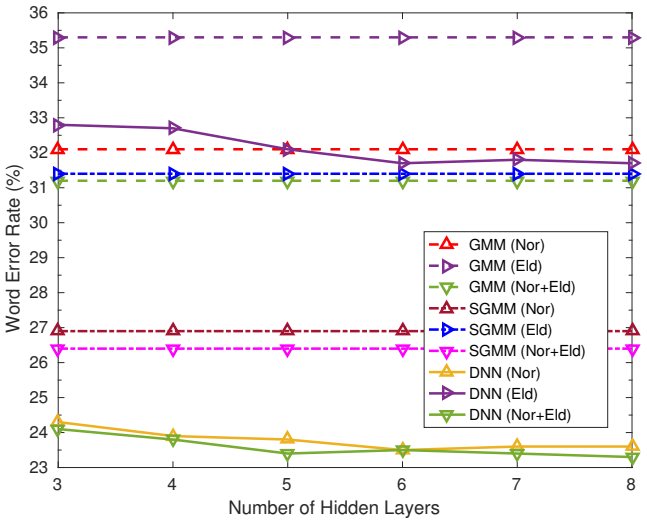

Figure 3: Word error rates obtained on Dutch dysarthric data.

The DNNs with 3 to 8 hidden layers each having 2048 sigmoid hidden units were trained on the FMLLR-adapted features. The DNN training was done by mini-batch Stochastic Gradient Descent with an initial learning rate of 0.008 and a minibatch size of 256. The time context size was 11 frames achieved by concatenating ±5 frames. A conventional trigram language model was trained on the transcriptions of the target sentence reading tasks.

4.2 Results

The word error rates (WER) provided by the different ASR systems trained on normal (Nor), elderly (Eld) and combined (Nor+Eld) Dutch speech are presented in Figure 3. The GMM(-HMM) trained on speaker adapted features of Nor, Eld and Nor+Eld data provides a WER of 32.1%, 35.3% and 31.2%, respectively. The acoustic models trained only on elderly speech perform worse than the systems trained on normal speech in all cases due to the limited amount of elderly training data. The recognition performance of the recognizers trained on normal speech is marginally improved by training the acoustic models on the combined data. SGMM-based models show a considerable improvement over the conventional GMMs. The SGMM recognizer system trained on the combined data yields a WER of 26.4%.

The number of DNN hidden layers are varied to investigate the impact of this parameter on capturing the increased variation in dysarthric speech. The best performance is obtained by training an 8-layered DNN system on the combined data, resulting in a WER of 23.3%. This is comparable to the WER of 23.4% obtained with the 5-layered system. Based on the trends followed by the curves of these systems, we can conclude that the recognition accuracy does not improve using DNNs with more than 6 hidden layers. This result is similar to the previous findings on the influence of hidden layers on normal speech recognition [16, 22].

5. DISCUSSION AND CONCLUSIONS

In this paper we have reported on research addressing the development of a speech therapy game that incorpo-

rates ASR technology to provide practice and feedback to dysarthric patients. Based on interviews and playtests (see subsection 3.1) with real users a game was developed that is in line with current views on the importance of social and collaborative elements in game design, as described by Gerling et al [13]. In addition, taking account of potential decrease in cognitive skills and short-term memory in dysarthric patients [18], game elements are gradually introduced, textual instructions are minimized and direct feedback as opposed to indirect feedback is favoured.

Initial ASR experiments (see subsection 4.2) conducted on a database containing Dutch dysarthric speech comparable to that which will be produced in the game revealed improved levels of recognition performance, as shown in Figure 3. These positive results indicate that employing ASR technology for providing feedback to dysarthric patients appears to be feasible. The recent developments in acoustic modeling are the main reason for this increase in dysarthric speech recognition performance, as has been observed with normal speech and other types of deviant speech, e.g. children speech [9, 20]. In addition, in the game more constrained tasks can be adopted by asking patients to answer questions with a limited number of correct responses, thus making it easier for the ASR to select the correct utterance. To summarize, the results presented in this paper are promising and pave the way for more focused experiments that employ on-task speech as soon as this becomes available.

6. ACKNOWLEDGMENTS

This research is funded by the NWO research grant with ref. no. 314-99-101 (CHASING). We would like to thank all members of the chasing team for their contribution: Marjoke Bakker, Lilian Beijer, Douwe-Sjoerd Boschman, Lodewijk Loos, Paulien Melis, Jurre Ongering, Toni Rietveld and Sabine Wildevuur.

7. REFERENCES

[1] S. Al Hashimi. The role of paralinguistic voice-control of interactive media in augmenting awareness of voice characteristics in the hearing-impaired. In *CHI 2007*, pages 2153–2158, 2007.

[2] L. J. Beijer. *E-learning based Speech Therapy (EST): Exploring the potentials of e-health for dysarthric speakers*. PhD thesis, Radboud University Nijmegen, Nov. 2012.

[3] L. J. Beijer, T. C. Rietveld, M. M. van Beers, R. M. Slangen, H. van den Heuvel, B. J. de Swart, and A. C. Geurts. E-learning-based speech therapy: a web application for speech training. *Telemed J E Health*, 16(2):177–180, 2010.

[4] H. T. Bunnell, D. Yarrington, and J. B. Polikoff. STAR: Articulation Training for Young Children. In *Proc. INTERSPEECH*, pages 85–88, 2000.

[5] B. M. Chen, D. J. Calder, and G. Mann. Computer-Based Multimedia Speech Training Tool For Dyspraxic Clients. In *Proc. of Speech Science & Technology*, volume 2, pages 504–509, 1994.

[6] G. Constantinescu, D. Theodoros, T. Russell, E. Ward, S. Wilson, and R. Wootton. Treating disordered speech and voice in Parkinson's disease online: a randomized controlled non-inferiority trial. *Int J Lang Commun Disord*, 46(1):1–16, 2011.

[7] K. R. Coventry, J. Clibbens, and M. Cooper. Specialist speech and language therapists' use and evaluation of visual speech aids. *Int J Lang Commun Disord*, 32(S3):315–323, 1997.

[8] C. Cucchiarini, J. Driesen, H. Van hamme, and E. Sanders. Recording Speech of Children, Non-Natives and Elderly People for HLT Applications: the JASMIN-CGN Corpus. In *Proc. LREC*, pages 1445–1450, 2008.

[9] G. Dahl, D. Yu, L. Deng, and A. Acero. Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition. *IEEE TASLP*, 20(1):30–42, 2012.

[10] J. R. Duffy. *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*. St Louis, MO, 1st edition, 1995.

[11] H. Fell, J. MacAuslan, J. Gong, C. Cress, and T. Salvo. visiBabble for Pre-speech Feedback. In *CHI '06 EA*, pages 767–772, 2006.

[12] S. Ferguson, A. Johnston, K. Ballard, C. T. Tan, and D. Perera-Schulz. Visual Feedback of Acoustic Data for Speech Therapy: Model and Design Parameters. In *Proc. Audio Mostly*, pages 135–140, 2012.

[13] K. M. Gerling, F. P. Schulte, J. Smeddinck, and M. Masuch. Game design for older adults: Effects of age-related changes on structural elements of digital games. In *Proc. ICEC*, pages 235–242, 2012.

[14] M. Glykas and P. Chytas. Technology assisted speech and language therapy. *Int J Med Inform*, 73(6):529 – 541, 2004.

[15] J. Hailpern, K. Karahalios, L. DeThorne, and J. Halle. Vocsyl: Visualizing Syllable Production for Children with ASD and Speech Delays. In *Proc. ASSETS*, pages 297–298, 2010.

[16] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Process Mag*, 29(6):82–97, 2012.

[17] S. E. Hutchins. SAY & SEE: Articulation Therapy Software. In *Proc. CAAPWD*, pages 37–40, 1992.

[18] R. D. Kent, J. F. Kent, J. Duffy, and G. Weismer. The dysarthrias: Speech-voice profiles, related dysfunctions, and neuropathology. *J Med Speech Lang Pathol*, 6(4):165–211, 1998.

[19] M. Krause, J. Smeddinck, and R. Meyer. A digital game to support voice treatment for Parkinson's disease. In *CHI '13 EA*, pages 445–450, 2013.

[20] H. Liao, G. Pundak, O. Siohan, M. K. Carroll, N. Coccaro, Q.-M. Jiang, T. N. Sainath, A. Senior, F. Beaufays, and M. Bacchian. Large Vocabulary Automatic Speech Recognition for Children. In *Proc. INTERSPEECH*, 2015.

[21] P. A. Mashima, D. P. Birkmire-Peters, M. J. Syms, M. R. Holtel, L. P. A. Burgess, and L. J. Peters. Telehealth: Voice Therapy Using Telecommunications Technology. *Am J Speech Lang Pathol*, 12(4):432–439, 2003.

[22] A.-R. Mohamed, G. E. Dahl, and G. Hinton. Acoustic Modeling Using Deep Belief Networks. *IEEE Trans Audio Speech Lang Processing*, 20(1):14–22, 2012.

[23] A. Morawej, A. T. Jackson, and R. D. McLeod. Fonetix: building virtual speech therapy practicum over the Internet. *Stud Health Technol Inform*, 64:253–261, 1999.

[24] N. Oostdijk. The Spoken Dutch Corpus: Overview and First Evaluation. In *Proc. LREC*, pages 886–894, 2000.

[25] M. Pietrowicz and K. G. Karahalios. Visualizing vocal expression. In *CHI 2014*, pages 1369–1374, 2014.

[26] D. Povey, L. Burget, M. Agarwal, P. Akyazi, F. Kai, A. Ghoshal, O. Glembek, N. Goel, M. Karafiát, A. Rastrow, R. C. Rose, P. Schwarz, and S. Thomas. The subspace Gaussian mixture model - A structured model for speech recognition. *Comput Speech Lang*, 25(2):404 – 439, 2011.

[27] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely. The Kaldi speech recognition toolkit. In *Proc. ASRU*, 2011.

[28] L. Ramig, S. Sapir, S. Countryman, A. Pawlas, C. O'Brien, M. Hoehn, and L. Thompson. Intensive voice treatment (LSVT®) for patients with parkinson's disease: a 2 year follow up. *J Neurol NeuroSur PS*, 71(4):493–498, 2001.

[29] W. R. Rodríguez, O. Saz, E. Lleida, C. Vaquero, and A. Escartín. Comunica - tools for speech and language therapy. In *Proc. WOCCI*, 2008.

[30] O. Saz, S.-C. Yin, E. Lleida, R. Rose, C. Vaquero, and W. R. Rodríguez. Tools and technologies for computer-aided speech and language therapy. *Speech Commun*, 51(10):948–967, 2009.

[31] H. Strik. ASR-based systems for language learning and therapy. In *Proc. of IS-ADEPT*, pages 9–14, 2012.

[32] C. T. Tan, A. Johnston, K. Ballard, S. Ferguson, and D. Perera-Schulz. sPeAK-MAN: towards popular gameplay for speech therapy. In *Proc. Australasian IE*, number 28, 2013.

[33] D. G. Theodoros. Telerehabilitation for service delivery in speech-language pathology. *J Telemed Telecare*, 14:221–224, 2008.

[34] E. van Leer and N. Porcaro. Pervasive diagnosis and rehabilitation of voice disorders: Current status and future directions. In *Proc. Future of Pervasive Health Workshop*. ACM, June 2016.

[35] K. Vicsi, P. Roach, A. Öster, Z. Kacic, P. Barczikay, A. Tantos, F. Csatári, Z. Bakcsi, and A. Sfakianaki. A multimedia, multilingual teaching and training system for children with speech disorders. *Int J Speech Tech*, 3:289–300, 2000.

[36] E. Yılmaz, M. Ganzeboom, L. Beijer, C. Cucchiarini, and H. Strik. A Dutch Dysarthric Speech Database for Individualized Speech Therapy Research. In *Proc. LREC*, 2016.