An ASR-Based Interactive Game for Speech Therapy

Mario Ganzeboom[1]*, Emre Yılmaz*[1]*, Catia Cucchiarini*[1] *and Helmer Strik*[1]

[1]CLS/CLST, Radboud University, Nijmegen, The Netherlands

{m.ganzeboom,e.yilmaz,c.cucchiarini,w.strik}@let.ru.nl

Abstract

The demand for intensive and costly speech therapy to patients impaired by communicative disorders can potentially be alleviated by developing computer-based systems that provide automatized speech therapy in the patient's home environment. In this paper we report on research aimed at developing such a system that combines serious gaming with automatic speech recognition (ASR) technology to provide computer-based therapy to dysarthric patients. The aim of the serious gaming environment is to increase the patients' motivation to practice, which tends to decrease over time with conventional speech therapy, as progress in dysarthric patients is often slow. Additionally, some speech exercises (e.g. drills) are not particularly motivating due to their repetitive nature. The ASR technology is aimed at providing feedback on speech quality during training to improve speech intelligibility. Different types of acoustic models were trained on normal speech of adults and elderly people, and tested on dysarthric speech. The results show that speaker-adaptive training and Deep Neural Networks (DNN)-based acoustic models substantially improve the performance of ASR in comparison to traditional GMM-HMM-based methods. In this specific case, the ASR-based game is developed to provide speech therapy to dysarthric patients, but this approach can be adapted for use in other types of communicative disorders.

Index Terms: communicative disorders, speech therapy, serious gaming, ASR.

1. Introduction

Among the problems that are likely to be associated with 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). Possible consequences of such diseases are communicative disorders. One of them is dysarthria, a motor speech disorder that affects speech intelligibility and causes communication problems [1]. Face-to-face speech therapy has proven beneficial for improving speech intelligibility in dysarthric patients, but to be effective therapy should be intensive [2, 3, 4, 5]. Owing to the increasing number of patients and the related high expenses, it may become difficult to provide intensive care in the future. As a result, attempts are being made at finding alternative, sustainable solutions that can guarantee the amount of care that is required for dysarthria patients in addition to or even without face-to-face sessions [6, 7]. In analogy to applications for pronunciation improvement in second language learning, [8, 9, 10], computer-based systems that employ ASR technology can be used to provide dysarthric patients with more robust and focused practice. A compounding problem however is that progress in these patients is slow, which is likely to reduce their motivation to practice. So one of the challenges is to develop systems that can motivate patients to get the necessary amount of practice. This can be achieved by resorting to games, which are known to increase motivation in learners and patients [11]. This aim is pursued in the CHASING project[1], in which a serious game employing ASR is being developed and evaluated to provide additional speech therapy to dysarthric patients.

In this paper we report on research that we conducted to develop and optimize this ASR-based game. Although we briefly refer to the process of game development and optimization, the emphasis is mainly on developing the ASR technology to be integrated in this game. The remainder of the paper is organized as follows. Section 2 briefly summarizes related work on game-based and ASR-based speech therapy. Section 3 presents the architecture of the ASR-based game. Section 4 describes the methodology adopted in experiments aimed at investigating how ASR can be improved to be incorporated in the game. Section 5 reports on the results of these experiments, while Section 6 presents a discussion of the results and the conclusions

2. Games and ASR for speech therapy

Neurological disorders like PD, stroke or TBI manifest more frequently at later ages (e.g. 55 or above), although these disorders sometimes may also occur at a younger age. Our research focuses on developing and evaluating an ASR-based serious game for providing speech therapy to elderly individuals with dysarthria, because these constitute the majority of the patients' group. Krause et al. [12] reported of work in this direction. They developed a game that challenged patients with PD to break glasses and vases by producing sufficiently loud and long /a/-phonemes. The game aimed to improve the reduced voice intensity of the patients that were reported to have a mild form of dysarthria. Speech processing algorithms sufficed to provide real-time feedback on the current and desired intensity only. An initial evaluation of the game was conducted with eight patients. Significant improvements of average peak voice loudness were observed in comparison with previously calibrated limits to those measured during game play.

Outside of academic research, similar types of games already existed: Dr. Speech[2] and VoxGames [3]. These games targeted children with varying speech disorders. RodrÃguez et al. [13] describe a set of small games named 'PreLingua'. These games were intended to assist the work in speech therapy focussing on deviations in phonatory related speech dimensions. The aim was to improve the use of voice activity, intensity, breathing, tone and vocalization of children with develop-

[1]See http://hstrik.ruhosting.nl/chasing/. Last retrieved on June 24, 2016.

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

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

mental disorders. Speech processing algorithms were utilized to analyse children's voices and control interactive elements in the games. Although the games were in use at a school for special education and were positively evaluated by a group of speech therapists, no evidence related to the efficacy of the games was reported.

Bunnell et al. [14] described work that included games and ASR technology. Their STAR system was intended for children with articulation disorders who practiced speech production starting with CV syllables progressing to words/phrases. Hidden Markov Models (HMMs) were trained on children's speech and evaluated on speech of children substituting /w/ for /r/. The results reported in their research showed that the log likelihoods from the HMMs correlated well with the perceptual ratings collected for utterances that contained substitutions, but poorly for correctly pronounced examples.

Vaquero et al. [15] introduced a set of small games named 'Vocaliza' as an addition to the previously described PreLingua. ASR technology was used to recognize the words children spoke while completing speech exercises. The novel addition was the utilization of ASR-based utterance verification (UV) technology to detect mispronunciations in a child's speech on a word level by calculating a confidence measure based on likelihood ratios. Similar to their research describing PreLingua, no results on performance evaluation were reported.

In a collaboration with Yin [16], they introduced mispronunciation detection at a phoneme level and obtained 6.7% absolute improvement in Equal Error Rate (EER) when replacing their baseline speaker-independent acoustic models with speaker-adapted models. In [17] they reported results on mispronunciation detection on both word and phoneme levels. They showed that their word-level pronunciation verification system in Vocaliza was rather unreliable for speech therapy due to a trade-off lowering the False Rejection Rate (FRR), at the cost of increasing the False Acceptance Rate (FAR). This limited frustration to the user, but at the same time accepted many mispronounced words. Their phoneme-level mispronunciation detection method did not show this trade-off and obtained considerable improvements in EER (i.e. equal FRR and FAR percentage). Additionally, further improvements in detecting mispronunciations were reported by fusing prior knowledge of the target word, target phoneme and its position in the word with the obtained posterior probabilities using MultiLayer Perceptron Neural Networks (NN-MLP).

Notable is also the study by Tan et al. [18] in which they combine word-level articulation exercises with popular gameplay (i.e. Pac-Man) employing an off-the-shelve ASR package. Feedback on the user's pronunciation was provided by triggering a predefined action in the game if a word was recognized and by displaying the recognized word and its corresponding confidence score at the top of the game screen. Evaluation took place using an informal play test with two children. Noteworthy observations are that the ASR package, not adapted to speech of children, frequently did not recognize their correct pronunciations of the target words. However, the children appeared to stay engaged and interested and continued playing. This may point to a certain level of tolerance for recognition errors, before a user gets frustrated.

As the previous paragraphs show, most research involving ASR-based games for speech therapy was aimed at disordered speech of children. In our research we developed a game aimed at elderly patients with dysarthric speech. To the best of our knowledge, this has not been done before. In addition to voice intensity in Krause et al. [12], our game also employs speech processing algorithms to provide real-time feedback on fundamental frequency (F0). We are currently researching strategies to also add feedback on pronunciation to the game. This strategy potentially consists of two phases that both employ ASR technology for pronunciation evaluation: utterance verification and pronunciation error detection. For the utterance verification phase we need to recognize the user's utterance. Mustafa et al. [19] provide an interesting overview of previous research on ASR for dysarthric speech. In section 4 we report on our initial recognition experiments that include dysarthric speech from our target group. Some previous research on pronunciation error detection in elderly dysarthric speech has been conducted [20, 21], but this addressed dysarthric speech due to different etiologies.

In the introduction we argued that serious games can increase patients' motivation for speech therapy. It is therefore important to limit potential sources of frustration in the game. Utterance verification and pronunciation error detection technology could be a source of such frustration, because it is not guaranteed that the patient's utterance is always recognized. In previously described research, patients potentially had to say the same utterance multiple times if it was not recognized by the system, before they were able to continue. This causes frustration to the patient and potentially lowers motivation. Elements in the gameplay should therefore not completely rely on the outcome of ASR technologies. A nice example is perhaps the game in Tan et al. [18]. The patient was only rewarded with a 'power up' if the associated word was recognized, but could still continue playing the game normally if this was not the case.

In the next section we outline our current game and briefly describe our ideas for integrating gameplay elements that do not fully rely on the outcomes of ASR technology.

3. The CHASING game for ASR-based speech therapy to dysarthric patients

In the CHASING project a serious game has been developed that Dutch-speaking dysarthric patients can play with their friends and relatives. The choice of the game was based on user tests in which several game concepts had been proposed and evaluated. An important aspect in this respect was whether the game should be a single player game or a multiplayer game. A single player game has the advantage that it can be played independently by a patient, without having to rely on other participants. On the other hand, multiplayer games are generally more engaging and motivating and are therefore likely to be played more frequently, which is of course very important for the therapy to be effective. The patients in our focus group showed a clear preference for multiplayer games and indicated that finding players would not be a problem as these could be their friends and relatives. Further tests with initial versions of the game revealed that additional considerations had to be made for the intended target group of elderly people who are no experienced gamers. For instance, it turned out that it was necessary to proceed more gradually both in introducing new game elements and in advancing to higher levels of difficulty. Moreover, the use of direct visual feedback was found easier to interpret as opposed to indirect feedback integrated into the gameplay. This is potentially due to diminished cognitive skills as an additional consequence of their neurological disorder. The game that was eventually selected and developed is called 'Schatzoekers' (i.e. 'Treasure hunters'). It is a two-player cooperative game in which players talk to each other through an audio con-

nection and have to help each other in finding the treasure and the key to open it. One player, the 'digger', can dig up the treasure on land and the other, the 'diver', dives in various rivers and canals in search for the key to open it. The locations of both treasure and key are marked on the map, but 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 (where the 'digger' should go). The players thus have to explain to each other where to go. This way, players are encouraged to keep speaking to each other to describe where they are on the map and giving directions where the other should go. Figure 1 shows the tablet set up for which the game was developed. Figure 2 displays a screenshot of the game.

Figure 1: The tablet set up displaying the start screen of the game.

Every map in the game is a different level. Levels of difficulty are influenced by the size and layout of the map, the complexity of street names and icons to describe one's location, the availability of an overview map and its level of detail. In the initial levels, your location is also visible to your co-player, in addition to the location of the item you need to find. That visibility is removed in later levels. Players talk to each other using the headset and get feedback on their loudness of voice and pitch from the game, especially when they are above or below specified thresholds. This is indicated by the horizontal green bar shown in Figure 2, which provides real-time feedback while the patient is speaking and shows a green, orange or red color when the loudness of the patient's speech is within, near or below the threshold, respectively. When the pitch is too high a notification slides down from underneath the bar instructing to 'speak loud and low'. The therapeutic goals of the game are to motivate dysarthric individuals in using continuous speech, and to speak up and maintain predefined levels of pitch and loudness.

In addition to feedback on loudness and pitch, our idea is to incorporate ASR technology in the game to be able to automatically provide more robust and focused feedback on speech quality. An initial idea to integrate this into the gameplay is to present the user with passphrases that have to be uttered, before the treasure is opened in the game. The user is always rewarded with treasure, but potentially depending on the level of speech quality determined by the ASR, the user may be rewarded with different kinds and/or amounts of treasure.

In the preparations for providing this type of feedback using ASR, we developed an ASR architecture that runs on a server in the cloud. Every time the game authenticates to this server, a separate ASR session is initialized which is only available to the

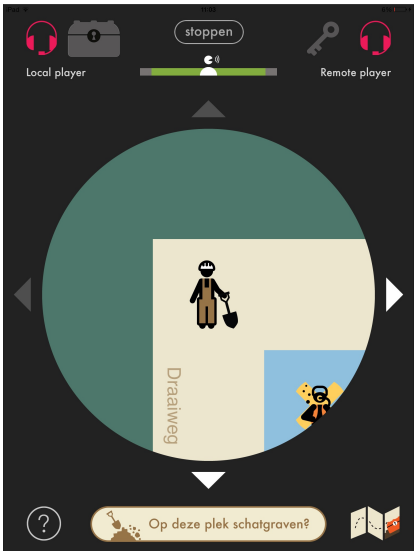

Figure 2: An in-game screenshot displaying the game from the perspective of the player 'digger'. In the partially blue square it can be observed that the, 'diver' already reached the correct location of the key.

user who requested it. The audio containing the player's speech is then continuously streamed to the server for offline analysis later on. As we have to handle privacy sensitive data, all communication with the server happens over secured connections.

We are currently developing and optimizing the ASR technology for the game and this work has to be done while the game is still being developed, improved and finalized. This means, among other things, that we cannot test the technology on the actual speech that will be produced in the game. A compounding problem in developing ASR applications for pathological speech is the limited availability of sufficient representative data. Since this was all anticipated, we started experimenting with already available speech data that can be considered representative for the type of speech that will have to be dealt with in the game (see section 4.1). Initial experiments were run to investigate to what extent ASR performance can be improved by speaker-independent Subspace Gaussian Mixture Model-Hidden Markov Models (SGMM-HMMs) and speaker-adaptive Deep Neural Networks (DNNs) in comparison to the traditional system using speaker-adaptive GMM-HMMs.

4. ASR experiments for the CHASING game

The ASR module in the CHASING game has to process dysarthric speech which is notoriously more difficult to recognize than normal speech. One of the obstacles in developing ASR technology that can handle dysarthric speech is the limited amount of dysarthric speech data available for training and testing the ASR algorithms. To partly circumvent this problem experiments were conducted in which maximum use was made of existing databases.

To investigate the baseline performance of the deep neural network-based acoustic models on dysarthric speech, we performed recognition experiments on a similar type of speech in Flemish which is a variety of the Dutch language spoken in Flanders. This choice is motivated by the availability of a prin-

cipled Flemish pathological speech database, namely the COPAS database [22], and the phonetic similarity between the two varieties Flemish and Dutch.

4.1. Speech databases

Within the framework of the CHASING project, a database of dysarthric speech is being collected [23], but this corpus is still relatively limited. For Dutch another, larger corpus of pathological speech has been collected [22], which also contains a considerable number of recordings of dysarthric speech. Although this database was compiled in Flanders and contains speech of patients who speak the Southern variety of Dutch (Flemish), it can be useful to investigate the baseline performance of the deep neural network-based acoustic models on dysarthric speech. First, the two varieties of Dutch spoken in the Netherlands and in Flanders are mutually intelligible and the most important phonological and phonetic differences are well known. Second, developing and testing the ASR on Flemish speech material makes it more feasible to adapt the CHASING game for Flemish patients at a later stage.

4.1.1. Training Data

Since the idea was to investigate the performance of the deep neural network-based acoustic models on Flemish dysarthric speech, Flemish speech data were used for training. These were obtained from the Flemish components of two Dutch and Flemish speech databases, i.e. CGN [24] and JASMIN-CGN [25]. The Flemish CGN component contains recordings of standard Flemish as spoken by adults in different regions of Flanders. The components with read speech, spontaneous conversations, interviews and discussions were used for training the acoustic models. The total duration of the normal Flemish speech (FLN) used in the recognition experiments is 186.5 hours. Additional speech material was taken from the Flemish component of the JASMIN-CGN corpus, which is an extension of the CGN database with speech of children non-natives and elderly people. The elderly speech component, with a total duration of approximately 5 hours, was employed in our experiments.

4.1.2. Testing Data

The COPAS pathological Flemish speech database [22] was used for testing the acoustic models trained on various speech types described in the previous section. The COPAS database has been collected within the framework of the SPACE project which was aimed at developing a reliable ASR-based speech assessment tool for pathological speech. This speech database contains recordings of 122 normal speakers as a control group and 197 speakers with speech disorders such as dysarthria, cleft, voice disorders, laryngectomy and glossectomy. The speech material includes not only word reading tasks, but also isolated sentence and short passage reading tasks.

The word reading tasks used in this paper is the Dutch Intelligibility Assessment (DIA) [26] material which contains 35 versions of 50 consonant-vowel-consonant (CVC) words and pseudowords organized in 3 subgroups. Moreover, we added all sentence reading tasks with annotations. These include 2 isolated sentence reading tasks (S1 and S2), 11 text passages (S) of reading difficulty levels AVI 7 and AVI 8 according to a system adopted in the Dutch language area that indicates reading difficulty based on text structure, vocabulary and length of words and sentences, and varies from AVI 1 up to AVI 9, and a phonetically balanced text known as Text Marloes (TM) [27].

For the recognition experiments, we classified the aforementioned material based on the type of speaker (normal vs. pathological) and speech material (word vs. sentence) resulting in 4 test sets. The speech segments in which the speaker does not utter the target word are discarded to be able to evaluate the recognizer errors only. There are 687 different words and 212 different sentences in the test data. The test set containing the word tasks uttered by normal speakers (WN) and speakers with disorders (WD) consists of 6154 and 8648 utterances with a total duration of 1.5 and 2 hours, respectively. The test set containing the sentence tasks uttered by normal speakers (SN) and speakers with disorders (SD) consists of 1918 (15,149) and 1034 (8287) sentences (words) with a total duration of 1.5 and 1 hour, respectively.

4.2. Implementation Details

The recognition experiments are performed using the Kaldi ASR toolkit [28]. The standard training recipe provided for multiple databases is applied to train a conventional context-dependent GMM-HMM on MFCC, LDA-MLLT and FMLLR-adapted features. Then, a system using an SGMM-based [29] acoustic model is also trained with a universal background model having 800 Gaussians and substate phone-specific vector size of 40. Providing the best performance among the aforementioned recognizers, this system is used to obtain the state alignments required for DNN training.

For DNN training, a standard feature extraction scheme is used by applying Hamming windowing with a frame length of 25 ms and frame shift of 10 ms. The DNNs with 6 hidden layers and 2048 sigmoid hidden units at each hidden layer were trained on the FMLLR-adapted features. The DNN training is 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 is 11 frames achieved by concatenating ± 5 frames. Unigram language models were trained on the target word transcriptions and used in the word recognition tasks. For the sentence recognition tasks, trigram language models were trained on the target sentence transcriptions.

5. Results and Discussion

We performed ASR experiments using the speech data described in Section 4.1. The recognition results obtained on the word and sentence tasks uttered by normal and pathological speakers from the COPAS database are presented in Table 1. For each column, the best results are marked in bold. In the context of the proposed serious game, sentence recognition is a more relevant task compared to isolated word recognition. For completeness, we present both word and sentence task results in this section.

Table 1: Word error rates in % obtained on the word and sentence COPAS test sets

Acoustic models	WordDys	WordNor	SentDys	SentNor
GMM+MFCC	76.2	55.0	37.3	13.3
GMM+LDA-MLLT	73.8	51.6	36.7	11.7
GMM+FMLLR	66.2	41.0	27.8	7.8
SGMM	59.2	34.0	**23.6**	5.7
DNN+FMLLR	**56.2**	**30.2**	**23.6**	**4.2**

The conventional GMM-HMM trained on Mel Frequency Cepstral Coefficients (MFCCs) provides a WER of 37.3% on

the dysarthric sentence utterances and a WER of 76.2% on the dysarthric word tasks. The WER difference between the normal and dysarthric speakers on the two tasks is larger than 20% for this system. The high WERs on the word tasks were due to the challenging recognition of one-syllable words and phonetically similar pseudowords. By using LDA-MLLT the WERs were reduced slightly as the second row in Table 1 shows.

Using discriminately trained features and including speaker information by applying speaker adaptive training (SAT) further reduced the WERs to 27.8% and 66.2%. Compared to GMM+LDA-MLLT, this is an absolute improvement of 8.9% and 7.6% on sentence and word tasks, respectively. The dysarthric speech recognizer benefits considerably from the speaker adaptive training.

Training SGMM-based acoustic models improves the recognition accuracy on the two tasks to a WER of 23.6% for sentence recognition and 59.2% on word recognition tasks. The DNN-based recognizer provides a similar performance with the SGMM-based recognizer on the sentence recognition task. An absolute improvement by 3% is obtained using DNNs on the word recognition task.

By using state-of-the-art DNN-based acoustic models it was possible to substantially lower the WERs, especially for the sentence task. In practice, we could lower the WERs even more, by using simpler tasks (exercises) with less complex language models. For instance, we could elicit speech in such a way that the number of possible correct answers is low, and then the ASR only has to determine whether one of these answers was spoken.

6. Conclusions

In this paper we have reported on our research aimed at developing an ASR-based game that can provide speech therapy to dysarthric patients with Dutch as their mother tongue. In particular, we have described experiments in which different types of acoustic models trained on normal speech were tested on dysarthric speech. The results show that speaker-adaptive training and Deep Neural Networks (DNN)-based acoustic models substantially improve the performance of ASR in comparison to traditional GMM-HMM-based methods.

Considering that the performance can further be improved by adopting more specific tasks in the game, as discussed in the previous section, we can conclude that the levels of accuracy obtained in these experiments bode well for the deployment of ASR in speech therapy applications. The experiments reported on in this paper were conducted on dysarthric speech of a close variety of Dutch, i.e. Flemish. Given that data sparsity is one of the major obstacles in developing ASR-based speech therapy applications, employing speech data of a closely related language variety is a possible way of approaching this problem. In the near future we intend to conduct similar experiments with dysarthric speech of the Northern variety of Dutch. However, since developing ASR-based speech therapy applications is very costly, it is important to know to what extent they are portable to other language varieties, so that more patients can profit from them.

Finally, we would like to underline that although in this specific case the ASR-based game has been developed to provide speech therapy to dysarthric patients, this approach can be easily adapted for use in other types of communicative disorders. To conclude, these results thus indicate that employing ASR technology for speech therapy to patients with communicative disorders is becoming more viable. In turn, this will allow them to get more intensive therapy, also in their home environment.

7. Acknowledgements

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.

8. References

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

[2] 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," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 71, no. 4, pp. 493–498, 2001.

[3] S. K. Bhogal, R. Teasell, and M. Speechley, "Intensity of aphasia therapy, impact on recovery," *Stroke*, vol. 34, no. 4, pp. 987–993, 2003.

[4] G. Kwakkel, "Impact of intensity of practice after stroke: Issues for consideration," *Disability and Rehabilitation*, vol. 28, no. 13-14, pp. 823–830, 2006.

[5] M. Rijntjes, K. Haevernick, A. Barzel, H. van den Bussche, G. Ketels, and C. Weiller, "Repeat therapy for chronic motor stroke: A pilot study for feasibility and efficacy," *Neurorehabilitation and Neural Repair*, vol. 23, no. 3, pp. 275–280, 2009.

[6] L. J. Beijer, A. C. M. Rietveld, M. B. Ruiter, and A. C. H. Geurts, "Preparing an e-learning-based speech therapy (est) efficacy study: Identifying suitable outcome measures to detect within-subject changes of speech intelligibility in dysarthric speakers," *Clinical Linguistics & Phonetics*, vol. 28, no. 12, pp. 927–950, 2014.

[7] H. Strik, "ASR-based systems for language learning and therapy," in *Proceedings of the International Symposium on Automatic Detection of Errors in Pronunciation Training*. Stockholm, Sweden: KTH, Computer Science and Communication, jun 2012, pp. 9–14.

[8] C. Cucchiarini, W. Nejjari, and H. Strik, "My pronunciation coach: Improving english pronunciation with an automatic coach that listens," *Language Learning in Higher Education*, vol. 1, no. 2, pp. 365–376, Nov. 2012.

[9] E. Krasnova and E. Bulgakova, *Proceedings of the 16th International Conference on Speech and Computer: SPECOM 2014*, 2014, ch. The Use of Speech Technology in Computer Assisted Language Learning Systems, pp. 459–466.

[10] B. P. de Vries, C. Cucchiarini, S. Bodnar, H. Strik, and R. van Hout, "Spoken grammar practice and feedback in an asr-based call system," *Computer Assisted Language Learning*, vol. 28, no. 6, pp. 550–576, 2015.

[11] P. M. Kato, S. W. Cole, A. S. Bradlyn, and B. H. Pollock, "A video game improves behavioral outcomes in adolescents and young adults with cancer: A randomized trial," *Pediatrics*, vol. 122, no. 2, pp. e305–e317, 2008.

[12] M. Krause, J. Smeddinck, and R. Meyer, "A digital game to support voice treatment for parkinson's disease," in *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '13. New York, NY, USA: ACM, 2013, pp. 445–450.

[13] W. R. Rodríguez, O. Saz, E. Lleida, C. Vaquero, and A. Escartín, "Comunica - tools for speech and language therapy," in *Proceedings of the 2008 Workshop on Children, Computer and Interaction*, 2008.

[14] H. T. Bunnell, D. Yarrington, and J. B. Polikoff, "Star: articulation training for young children," in *Sixth International Conference on Spoken Language Processing, ICSLP / INTERSPEECH*, Nov 2000, pp. 85–88.

[15] C. Vaquero, O. Saz, E. Lleida, J. M. Marcos, and C. Canalís, "Vocaliza: An application for computer-aided speech therapy in spanish language," in *IV Jornadas en Tecnologia del Habla*. Zaragoza, Spain: Grupo de tecnologias de las communicaciones, Universidad de Zaragoza, Nov. 2006, pp. 321–326.

[16] S.-C. Yin, R. C. Rose, O. Saz, and E. Lleida, "Verifying pronunciation accuracy from speakers with neuromuscular disorders," in *INTERSPEECH*. ISCA, Sept 2008, pp. 2218–2221.

[17] 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 Communication*, vol. 51, no. 10, pp. 948–967, 2009.

[18] C. T. Tan, A. Johnston, K. Ballard, S. Ferguson, and D. Perera-Schulz, "speak-man: towards popular gameplay for speech therapy," in *Proceedings of The 9th Australasian Conference on Interactive Entertainment: Matters of Life and Death*, no. 28, 2013.

[19] M. B. Mustafa, F. Rosdi, S. S. Salim, and M. U. Mughal, "Exploring the influence of general and specific factors on the recognition accuracy of an asr system for dysarthric speaker," *Expert Systems with Applications*, vol. 42, no. 8, pp. 3924 – 3932, 2015.

[20] Z. A. Benselama, M. Guerti, and M. A. Bencherif, "Arabic speech pathology therapy computer aided system," *Journal of Computer Science*, vol. 3, no. 9, pp. 685–692, 2007.

[21] T. Pellegrini, L. Fontan, J. Mauclair, J. Farinas, C. Alazard-Guiu, M. Robert, and P. Gatignol, "Automatic assessment of speech capability loss in disordered speech," *ACM Transactions on Accessible Computing*, vol. 6, no. 3, pp. 8:1–8:14, May 2015.

[22] C. Middag, "Automatic analysis of pathological speech," Ph.D. dissertation, Ghent University, Belgium, 2012.

[23] E. Yilmaz, M. Ganzeboom, L. Beijer, C. Cucchiarini, and H. Strik, "A dutch dysarthric speech database for individualized speech therapy research," in *Proc. LREC*, may 2016, pp. 792–795.

[24] N. Oostdijk, "The spoken Dutch corpus: Overview and first evaluation," in *Proc. LREC*. LREC, 2000, pp. 886–894.

[25] 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*, May 2008, pp. 1445–1450.

[26] M. De Bodt, C. Guns, and G. Van Nuffelen, "NSVO: handleiding," Vlaamse Vereniging voor Logopedie: Herentals, Tech. Rep., 2006.

[27] J. Van de Weijer and I. Slis, "Nasaliteitsmeting met de nasometer," *Logopedie en Foniatrie*, vol. 63, pp. 97–101, 1991.

[28] 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*, Dec. 2011.

[29] D. Povey, L. Burget, M. Agarwal, P. Akyazi, K. Feng, A. Ghoshal, O. Glembek, N. K. Goel, M. Karafiát, A. Rastrow, R. C. Rose, P. Schwarz, and S. Thomas, "The subspace gaussian mixture model - A structured model for speech recognition," *Computer Speech & Language*, vol. 25, no. 2, pp. 404–439, 2011.