IberSPEECH 2021
24-25 March 2021, Valladolid, Spain

Prosody training of people with Down syndrome using an educational video game

Mario Corrales-Astorgano[1]

[1]Department of Computer Science, University of Valladolid, Spain
mcorrales@infor.uva.es

Abstract

The speech of people with Down syndrome presents multiple disorders affecting the different components of language (syntax, semantics, phonology and pragmatics). In particular, prosody is also affected, conditioning their personal development and their social integration. Due to these difficulties, prosody training is fundamental in their speech therapy. One aim of this work is the definition of a video game focused on training some language skills, specifically those ones related with prosody. Other aim is the creation of an analysis system to evaluate the prosodic quality of users' utterances. The results show that the video game is useful to keep the attention of the players with Down syndrome during all game session. In addition, some statistically significant differences are found between people with Down syndrome and people without intellectual disabilities in the frequency, energy, temporal and spectral domain. The accuracy of identifying a recording as produced by a person with Down syndrome or by a person without intellectual disabilities is up to 95%. Finally, an accuracy of 79.3% is achieved in the task of predicting the prosodic expert evaluation using an automatic classifier trained with the acoustic features extracted from the recordings of people with Down syndrome.

1. Introduction

Some people with Down syndrome (DS) have problems in their social relationships due to their communicative problems [1, 2, 3]. Speech in general [4] and prosody in particular [5] are affected, producing problems in their communicative skills. There are few works in the literature that have analyzed the speech of people with Down syndrome using a corpus comparative approach [4]. The majority of the studies presented in the literature have used a perceptual test approach, while there are few studies based on comparing or analyzing acoustic features extracted from the recordings of the corpus. The recording of a speech corpus of people with Down syndrome is a hard task, because these people present cognitive problems like short time memory problems or attention deficits, among others [1]. In addition, characterizing prosodic impairments in populations with developmental disorders is a hard task [6]. However, prosody assessment procedures appropriate for using with individuals with intellectual and/or developmental disabilities need to be employed, with the aim of being useful for speech therapists in their therapy with people with Down syndrome.

The potential of games to improve motivation and engagement in education has been examined [7]. However, there are few of them focused on speech training of people with Down syndrome [4]. Although there are some tools described in the literature [8], the cognitive problems of people with Down syndrome difficult them to use these tools.

In this work, the main objective was the development of an educational video game focused on improving the communication skills of people with Down syndrome, specifically prosodic skills. To reach this objective, the main communication problems of this population had been analyzed with the aim of developing training activities focused on improving these problems. To improve the motivation of the players, it was important to define the scenarios and the narrative where the training activities were included. The analysis of how the cognitive problems of people with Down syndrome could affect the interaction with the video game had been studied to design an effective interface. In addition, the automatic evaluation of the audios recorded in the training activities could be useful to enhance the autonomous use of the video game by the players with Down syndrome. This was a thesis by compendium of publications, so each of the 3 papers was focused on different objectives of the thesis. The video game design and evaluation was analyzed in [9], the study of differences between acoustic features of the speech of people with Down syndrome and the speech of people without intellectual disabilities was developed in [10] and the automatic evaluation of the recordings of the video game was presented in [11].

The structure of the article is as follows. Section 2 reviews related works from the state of the art. Section 3 describes the characteristics of the video game. Section 4 describes a summary of the different methodologies followed to reach the thesis objectives. Section 5 shows a summary of the results obtained in this thesis related with the video game evaluation, the analysis of the differences between acoustic features of the speech of people with Down syndrome and the speech of people without intellectual disabilities and a first approach to the automatic evaluation of the video game recordings. Finally, section 6 describes the conclusions of this work.

2. Previous work

The are mainly two approaches that have been followed to analyze the prosody problems of people with Down syndrome: acoustic analysis and perceptual analysis. In both cases, the age of the population selected for the study seems to be important for the results obtained, due to the physiological differences between children and adults.

Concerning the acoustic approach, [12, 13, 14] found significantly higher F0 values in adults with Down syndrome as compared to adults with typical development (TD). In addition, [12] found lower jitter (frequency perturbations) in adult speakers with Down syndrome than in TD adults. As for energy, [14] found significantly lower energy values in adults with Down syndrome than in TD people. Moreover, [15] concluded that adults with Down syndrome had poor control over energy in stressed versus unstressed vowels. Finally, temporal domain results depend on the unit of analysis employed. [15] found that people with cognitive disorders presented an excessive variability in vowel duration, while [13] and [16] reported longer du-

10.21437/IberSPEECH.2021-37

rations of vowels in adults with Down syndrome than in TD adults. [14] discovered a lower duration of words in male adults with Down syndrome than male TD adults. Moreover, people with Down syndrome present some disfluency problems. Although disfluency (stuttering or cluttering) has not been demonstrated as a universal characteristic of Down syndrome, it is a common problem of this population [17, 18, 19]. These disfluencies can affect the speech rhythm of people with Down syndrome. On the other hand, [20] indicated that children with Down syndrome had lower F0 than children without intellectual disabilities. [21] found higher jitter in children with Down syndrome than children without intellectual disabilities. In terms of energy, [21] indicated higher shimmer in children with Down syndrome than in children without intellectual disabilities. Perceptual studies show mixed results. [21] described the voice of children with Down syndrome as being statistically different from the voice of children without intellectual disabilities in five speech problems: grade, roughness, breathiness, asthenic speech and strained speech. [22] judged the voice quality of adults with Down syndrome as hoarse. In addition, [23] noted discrepancies between perceptual judgments of pitch level and acoustic measures of F0.

Concerning educational video games, there are some studies that show the efficiency of ICT and video games in the cognitive rehabilitation and teaching of people with intellectual disability: improvement in choice reaction time [24], stimulating cognitive abilities of children [25], and independent decision making [26]. Furthermore, other games are focused on therapy for children with speech disorders [27]. Reviews of educational software for people with DS can be found in [28].

Automatic assessment of pathological speech has also been researched, but, in general, the studies on the topic are related to specific aspects and populations. Some works focus on the speech intelligibility of people with aphasia [29] or speech intelligibility in pathological voices [30]. Others try to identify speech disorders in children with cleft lip and palate [31] or to predict automatically some dysarthric speech evaluation metrics, such as intelligibility, severity and articulation impairment [32]. All these works include a subjective evaluation carried out by experts as a reference to train the classification systems.

3. Game description

The video game, named as PRADIA, has the structure of a graphic adventure game, including conversations with characters, getting and using items and navigating through scenarios (Figure 1). Players have to use the mouse to interact with the elements of the game. Players go through different scenarios where they have to do some actions, like solving an activity or using an item. Some activities focus on lexical-semantic comprehension and on improving of prosodic perception in specific contexts. Other activities focus on oral production, so the player is encouraged by the game to train his speech, keeping in mind prosodic aspects like intonation, expression of emotions or syllabic emphasis. In the activities, the player is introduced by the game into different conversations with game characters, where the player has to choose between different options to continue the dialogue (Comprehension activities) or to record some sentences related with the dialogue context (Production activities), depending on the activity. Finally, there are other activities that were included to add variety to the game and to train other skills not directly related to speech training (Visual activities).

During the game session, information about user interaction is stored, as well as the audio recordings of the production activities. This information was used to compare the interaction between people with Down syndrome and people without intellectual disabilities. Additionally, the audio recordings increase the speech corpus. This user interaction log has information about game and activities duration, the attempts to complete a task, number of mouse clicks or the helps showed to the user.

Figure 1: *Production activity inside a game scenario*

4. Methodology

Figure 2 shows the methodology followed to design, develop and evaluate the video game. This is a user-centered design methodology, where Down syndrome experts, therapist and final users participated in the game development. After the development, the game was evaluated by people with Down syndrome taken into account the five usability aspects defined in [33] (easy to learn, effective, efficient, engaging and error tolerant). These evaluations were carried out with 14 users with Down syndrome (10 boys and 4 girls, chronological age between 13 and 39 years), 10 children without intellectual disabilities and 10 adults without intellectual disabilities. To analyze the differences in the game interaction among groups, objective and subjective evaluation methods were combined. On the one hand, the game itself recorded data about the interaction between the player and the game. On the other hand, at the end of the game session, evaluators gave a questionnaire on general aspects of usability to the player; this was complemented by observations that evaluators collected during the test. In addition, at the end of the questionnaire, the opinions of speech therapists or teachers who are dedicated to special education were collected. As the video game records the voice of the players, we used these recordings in a perception test, to check whether the players' pronunciation improves as they use the video game.

Figure 3 shows the experimental methodology followed to identify the differences of the acoustic and prosodic features of people with Down syndrome in comparison with people without intellectual disabilities. Firstly, the speech corpus recorded by people with Down syndrome and by typically developing people was gathered (349 recordings of 18 speakers with Down syndrome and 250 recordings of 22 speakers without intellectual disabilities). Secondly, acoustic features were extracted from all the recordings of each corpus using the openSmile software [34] and the feature set GeMAPS [35]. Thirdly, a non-parametric statistical test to analyze the differences between groups was carried out. Fourthly, the automatic classification experiment was carried out, in which the features with signifi-

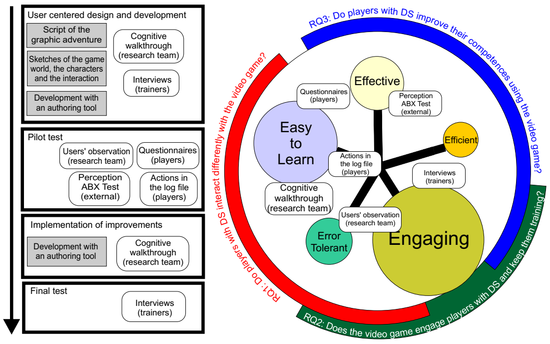

Figure 2: *Evaluation strategy. The circles represent the usability aspects (the size reflects the importance of the aspect), the rectangles represent the development and evaluation phases, and the rounded rectangles are the evaluation activities with the participants in brackets.*

cant differences were used to train three binary classifiers. The performance of these classifiers was obtained using 10-folds cross validation, where 90% of the data was used to train the classifier and 10% was used to test the classifier, repeating this process 10 times. To analyze the performance of the classification, we used the classification rate. The unweighted average recall (UAR) was also used. This metric is the mean of sensitivity (recall of positive instances) and specificity (recall of negative instances). UAR was chosen as the classification metric because it equally weights each class regardless of its number of samples, so it represents more precisely the accuracy of a classification test using unbalanced data. Finally, in order to evaluate the impact of prosody in the perception of the listeners, prosody transfer techniques were used. These techniques consist of transfers, phoneme by phoneme, of the pitch, energy and duration from one audio to another. Therefore, the new audio file contains the original utterance but with the prosody transferred from another utterance. This process was applied to some audios of people with Down syndrome and other audios from people without intellectual disabilities. Afterwards, some people without specific knowledge in prosody evaluated each modified recording using a 5-point Likert scale, indicating the security degree in which they identify each recording as belonging to a person with an intellectual disability.

Figure 4 describes the experimental procedure followed to analyze the automatic evaluation of the recordings of the video game. Three corpus of the recordings of the video game but gathered in different times were used. The first corpus contained recordings of 5 speakers with Down syndrome (605 recordings). The second corpus contained recordings of other 5 speakers with Down syndrome. And the third corpus contained recordings of 13 speakers with Down syndrome obtained with a previous version of the video game. These recordings were evaluated by a prosody expert following the categories of intonational phonology (intonation, accent and prosodic organization). The evaluation was binary, Right (R) or Wrong (W), depending of the quality of the recording. In addition, some recordings were evaluated by the therapist who was next to the players during the game sessions, using a 3-level scoring. If the evaluation was Cont.R (Continue with right result) or Cont. (Continue but the oral activity could be better), the video game advanced to the next activity. If the evaluation was Rep. (Repeat), the game offered a new attempt in which the player had to repeat the activity. With the aim of automatically predicting

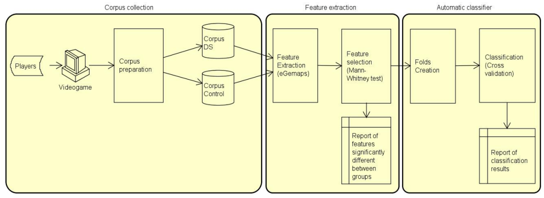

Figure 3: *Scheme of the experimental procedure to compare the speech of people with Down syndrome and the speech of people without intellectual disabilities.*

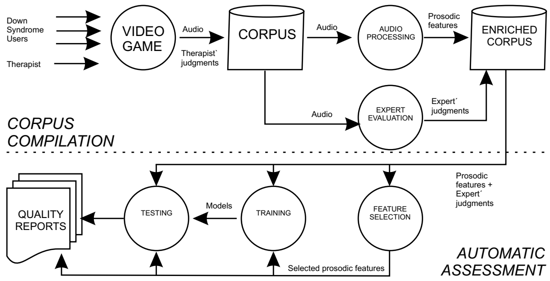

Figure 4: *Experimental procedure scheme followed to analyze the automatic evaluation of the video game recordings.*

the prosody expert evaluations, the same set of features used in the previous experiment (GeMAPS) was extracted from each recording. These features together with the prosody expert evaluations were used to train and test 3 classifiers using 10-folds cross validation. Finally, the automatic and human based evaluations of 5 speakers were compared to analyze the impact of speakers with Down syndrome heterogeneity on classification results.

5. Results and discussion

Related with the design and evaluation of the video game, the results showed that people with Down syndrome had more errors and needed more time to complete the activities (production and comprehension) than people without intellectual disabilities. The data gathered during the game sessions could be useful to detect specific prosodic problems of each player. This data can be used by speech therapists to train the skills related with these activities with the aim of improving them. In addition, players with Down syndrome showed a high motivation to complete the activities of the video game. This is an important result because people with Down syndrome present cognitive limitations that difficult them doing training tasks. The inclusion of the training activities inside a narrative helped to improve the motivation of the players. The questionnaires done by the speech therapists after the game sessions strengthened this results. On the other hand, the video game was useful to gather a speech corpus focused on the prosody of people with Down syndrome.

Table 1 shows the results related with identifying the speaker group (Down syndrome or typical speakers) of each recording of the corpus, using 3 automatic classifiers. The acoustic features related with fundamental frequency, energy and rhythm selected alone show worse performance results than spectral ones, but the combination of these 3 type of features obtained better performance results than the spectral alone. The spectral features are usually related with particularities of the

Table 1: *Classification results for identifying the group of the speaker. Classification rate (C.Rate) and UAR using different feature sets and different classifiers are reported. The features used are those with significant differences between TD and DS groups. The classifiers are decision tree (DT), support vector machine (SVM) and multilayer perceptron (MLP). # is the number of input features in each set.*

		SVM		MLP		DT	
Set	**#**	**C. Rate**	**UAR**	**C. Rate**	**UAR**	**C. Rate**	**UAR**
Frequency	9	62.67	0.61	64.33	0.64	60.17	0.60
Energy	9	79.33	0.78	76	0.76	72.50	0.71
Temporal	9	76.83	0.76	77.83	0.78	74.33	0.75
Frequency+Energy+Temporal	27	90	0.9	91.83	0.91	82	0.82
Spectral	34	87.33	0.87	87.33	0.87	84.33	0.84
All	61	94.17	0.94	95.17	0.95	86.50	0.87

Table 2: *Percentage of coincidence between therapist decision, classifier and prosody expert per speaker. Concerning the classifier, R represents the utterances classified as Right by the classifier and W represents the utterances classified as Wrong by the classifier. Each row percentage is relative to the number of each type of utterances of prosody expert evaluation.*

		Expert judgment		**Classified as**		**Therapist decision**		
Speaker	**#Total utt**	**type**	**#utt**	**R**	**W**	**Cont.R**	**Cont.**	**Rep.**
S01	120	R	87	83.91%	16.09%	68.97%	24.14%	6.90%
		W	33	57.58%	42.42%	30.30%	36.36%	33.33%
S02	106	R	81	87.65%	12.35%	85.19%	14.81%	0.00%
		W	25	28.00%	72.00%	84.00%	16.00%	0.00%
S03	97	R	78	97.44%	2.56%	94.87%	3.85%	1.28%
		W	19	73.68%	26.32%	100.0%	0.00%	0.00%
S04	131	R	75	94.57%	5.33%	21.33%	44.00%	34.67%
		W	56	41.07%	58.93%	5.36%	32.14%	62.5%
S05	151	R	77	87.01%	12.99%	20.78%	50.65%	28.57%
		W	74	29.73%	70.27%	6.76%	20.27%	72.97%
Total	605	R	398	89.96%	10.05%	59.06%	27.14%	13.84%
		W	207	41.06%	58.94%	28.05%	23.72%	48.31%

phonological system of people with Down syndrome, so the training of the use of these features is complicated or impossible. However, the use of the features related with fundamental frequency, energy and rhythm can be improved by therapy. This improving can be reflected on a better use of intonation, accent or rhythm in the speech of people with Down syndrome. In addition, the results of the perceptual test done to evaluate the prosody transferred recordings show the importance of prosody to identify a voice as atypical. The recordings of people without intellectual disabilities with the transferred prosody (frequency, energy and duration) of the recordings of people with Down syndrome were identified mainly as atypical speech as well as the recordings of people with Down syndrome with the transferred prosody of the recordings of people without intellectual disabilities were identified mainly as typical speech.

The results of the experiment to automatically evaluate the prosodic quality of the video game recordings show the impact of the heterogeneity presented by people with Down syndrome in this automatic evaluation (Table 2). In short, agreement between the prosodic expert and the therapist depends on the speaker's developmental levels and the type of sentence produced (right or wrong). In addition, differences in the evaluation context can also explain raters' disagreements. Thus, while the expert only based her decisions on intonational criteria, the therapist also took into consideration the progress of the player while playing the video game. In doing so, avoiding frustration was a priority; therefore, levels of frustration tolerance and number of failures influenced the therapist's decisions. In the video game, it is very important to avoid evaluating as wrong a correct utterance; otherwise, frustration may arise. Bearing in mind that the video game aims to engage and motivate the users, the percentage of false negatives must be as low as possible. These results show that only 10.1% of the samples evaluated as Right by the expert are classified as Wrong by the classifier.

6. Conclusions and future work

Concerning the video game design, the narrative and the training activities were developed in collaboration with experts in intellectual disabilities, experts in language deficits, the therapists of the centers of special education and the final users. This process is important because it allowed the detection of possible errors in the design before the development phase. The evaluation carried out by the people with Down syndrome allowed the extraction of some key aspects for the development of educational tools focused on people with Down syndrome. The realization of the training activities is the main objective when a educational video game is developed, so engagement is a very important usability aspect that has to be taken into account to motivate the players.

In addition, the possibility of recording a speech corpus using the video game is relevant specially when the target population has some cognitive limitation. The corpus gathered allowed the comparison between the prosody of people with Down syndrome and the prosody of people without intellectual disabilities using the acoustic features extracted from the recordings. The experiments carried out using automatic classifiers to identify the speaker group that produced a recording obtained high performance results using features related with fundamental frequency, energy, temporal and spectral domains. On the other hand, the perceptual experiment carried out using the recordings created by the transferred prosody algorithm showed the importance of prosody to identify speech as atypical. This result is important because the prosody training can help people with Down syndrome to improve their integration in society.

Furthermore, the heterogeneity presented by people with Down syndrome affected the automatic evaluation of the recordings quality and the perceptual evaluation done by experts. The concordance values among the therapist, the prosody expert and the automatic classifiers varied depending of the cognitive level and speech quality of each speaker. These results suggest that the automatic classifiers focused on evaluating the prosody quality of people with Down syndrome have to take into account the heterogeneity of this population with the aim of obtaining better performance results.

As future work, the adaptation of the training activities to each speaker profile can improve the learning process. In addition, the generation of a report of the results of each speaker can help therapists to focus on the specific problems of each speaker with the aim of doing a personalized training.

7. Acknowledgements

I want to thank my mentors David Escudero Mancebo and César González Ferreras for their support in the realization of this thesis. This thesis was defended on September 2019. The full paper thesis is available on http://uvadoc.uva.es/handle/10324/38470.

8. References

[1] R. S. Chapman, "Language development in children and adolescents with Down syndrome," *Mental Retardation and Developmental Disabilities Research Reviews*, vol. 3, no. 4, pp. 307–312, 1997.

[2] J. Cleland, S. Wood, W. Hardcastle, J. Wishart, and C. Timmins,

"Relationship between speech, oromotor, language and cognitive abilities in children with Down's syndrome," *International journal of language & communication disorders*, vol. 45, no. 1, pp. 83–95, 2010.

[3] G. E. Martin, J. Klusek, B. Estigarribia, and J. E. Roberts, "Language characteristics of individuals with Down syndrome," *Topics in Language Disorders*, vol. 29, no. 2, p. 112, 2009.

[4] R. D. Kent and H. K. Vorperian, "Speech impairment in Down syndrome: a review," *Journal of Speech, Language, and Hearing Research*, vol. 56, no. 1, pp. 178–210, 2013.

[5] V. Stojanovik, "Prosodic deficits in children with Down syndrome," *Journal of Neurolinguistics*, vol. 24, no. 2, pp. 145–155, 2011.

[6] S. J. Peppé, "Why is prosody in speech-language pathology so difficult?" *International Journal of Speech-Language Pathology*, vol. 11, no. 4, pp. 258–271, 2009.

[7] A. McFarlane, A. Sparrowhawk, and Y. Heald, *Report on the educational use of games*. TEEM (Teachers evaluating educational multimedia), Cambridge, 2002.

[8] O. Saz, S. 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.

[9] C. González-Ferreras, D. Escudero-Mancebo, M. Corrales-Astorgano, L. Aguilar-Cuevas, and V. Flores-Lucas, "Engaging adolescents with Down syndrome in an educational video game," *International Journal of Human–Computer Interaction*, pp. 1–20, 2017.

[10] M. Corrales-Astorgano, D. Escudero-Mancebo, and C. González-Ferreras, "Acoustic characterization and perceptual analysis of the relative importance of prosody in speech of people with Down syndrome," *Speech Communication*, vol. 99, pp. 90– 100, 2018.

[11] M. Corrales-Astorgano, P. Martínez-Castilla, D. Escudero-Mancebo, L. Aguilar, C. González-Ferreras, and V. Cardeñoso-Payo, "Automatic assessment of prosodic quality in down syndrome: Analysis of the impact of speaker heterogeneity," *Applied Sciences*, vol. 9, no. 7, p. 1440, 2019.

[12] M. T. Lee, J. Thorpe, and J. Verhoeven, "Intonation and phonation in young adults with Down syndrome," *Journal of Voice*, vol. 23, no. 1, pp. 82–87, 2009.

[13] A. Rochet-Capellan and M. Dohen, "Acoustic characterisation of vowel production by young adults with Down syndrome," in *18th International Congress of Phonetic Sciences (ICPhS 2015)*, Glasgow, United Kingdom, Aug. 2015. [Online]. Available: https://hal.univ-grenoble-alpes.fr/hal-01240561

[14] G. Albertini, S. Bonassi, V. Dall'Armi, I. Giachetti, S. Giaquinto, and M. Mignano, "Spectral analysis of the voice in Down syndrome," *Research in developmental disabilities*, vol. 31, no. 5, pp. 995–1001, 2010.

[15] O. Saz, J. Simón, W. Rodríguez, E. Lleida, C. Vaquero *et al.*, "Analysis of acoustic features in speakers with cognitive disorders and speech impairments," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, p. 1, 2009.

[16] K. Bunton and M. Leddy, "An evaluation of articulatory working space area in vowel production of adults with Down syndrome," *Clinical linguistics & phonetics*, vol. 25, no. 4, pp. 321–334, 2011.

[17] J. Van Borsel and A. Vandermeulen, "Cluttering in Down syndrome," *Folia Phoniatrica et Logopaedica*, vol. 60, no. 6, pp. 312–317, 2008.

[18] D. Devenny and W. Silverman, "Speech dysfluency and manual specialization in Down's syndrome," *Journal of Intellectual Disability Research*, vol. 34, no. 3, pp. 253–260, 1990.

[19] K. Eggers and S. Van Eerdenbrugh, "Speech disfluencies in children with down syndrome," *Journal of Communication Disorders*, vol. 71, pp. 72–84, 2018.

[20] L. Zampini, M. Fasolo, M. Spinelli, P. Zanchi, C. Suttora, and N. Salerni, "Prosodic skills in children with Down syndrome and in typically developing children," *International Journal of Language & Communication Disorders*, vol. 51, no. 1, pp. 74–83, 2016.

[21] C. P. Moura, L. M. Cunha, H. Vilarinho, M. J. Cunha, D. Freitas, M. Palha, S. M. Pueschel, and M. Pais-Clemente, "Voice parameters in children with Down syndrome," *Journal of Voice*, vol. 22, no. 1, pp. 34–42, 2008.

[22] M. Moran and H. Gilbert, "Selected acoustic characteristics and listener judgments of the voice of down syndrome adults," *American journal of mental deficiency*, vol. 86, no. 5, p. 553—556, March 1982.

[23] R. Rodger, "Voice quality of children and young people with Down's Syndrome and its impact on listener judgement," Ph.D. dissertation, Queen Margaret University, 2009.

[24] P. Standen, N. Anderton, R. Karsandas, S. Battersby, and D. Brown, "An evaluation of the use of a computer game in improving the choice reaction time of adults with intellectual disabilities," *Journal of Assistive Technologies*, vol. 3, no. 4, pp. 4–11, 2009.

[25] A. Brandão, L. Brandão, G. Nascimento, B. Moreira, C. N. Vasconcelos, and E. Clua, "Jecripe: stimulating cognitive abilities of children with Down syndrome in pre-scholar age using a game approach," in *Proceedings of the 7th International Conference on Advances in Computer Entertainment Technology*. ACM, 2010, pp. 15–18.

[26] P. Standen, F. Rees, and D. Brown, "Effect of playing computer games on decision making in people with intellectual disabilities," *Journal of Assistive Technologies*, vol. 3, no. 2, pp. 4–12, 2009.

[27] M. Cagatay, P. Ege, G. Tokdemir, and N. E. Cagiltay, "A serious game for speech disorder children therapy," in *Health Informatics and Bioinformatics (HIBIT), 2012 7th International Symposium on*. IEEE, 2012, pp. 18–23.

[28] B. Black, "Educational software for children with down syndrome-an update," *Down Syndrome News and Update*, vol. 6, no. 2, pp. 66–68, 2006.

[29] D. Le, K. Licata, C. Persad, and E. M. Provost, "Automatic assessment of speech intelligibility for individuals with aphasia," *IEEE/ACM transactions on audio, speech, and language processing*, vol. 24, no. 11, pp. 2187–2199, 2016.

[30] A. Maier, T. Haderlein, U. Eysholdt, F. Rosanowski, A. Batliner, M. Schuster, and E. Nöth, "Peaks–a system for the automatic evaluation of voice and speech disorders," *Speech Communication*, vol. 51, no. 5, pp. 425–437, 2009.

[31] A. K. Maier, F. Hönig, C. Hacker, M. Schuster, and E. Nöth, "Automatic evaluation of characteristic speech disorders in children with cleft lip and palate," in *INTERSPEECH 2008, 9th Annual Conference of the International Speech Communication Association, Brisbane, Australia, September 22-26, 2008*. ISCA, 2008, pp. 1757–1760.

[32] I. Laaridh, W. B. Kheder, C. Fredouille, and C. Meunier, "Automatic prediction of speech evaluation metrics for dysarthric speech," in *Proc. Interspeech 2017*, 2017, pp. 1834–1838.

[33] W. Quesenbery, "The five dimensions of usability," *Content and complexity: Information design in technical communication*, pp. 81–102, 2003.

[34] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in opensmile, the Munich open-source multimedia feature extractor," in *Proceedings of the 21st ACM international conference on Multimedia*. ACM, 2013, pp. 835–838.

[35] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. André, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan *et al.*, "The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing," *IEEE Transactions on Affective Computing*, vol. 7, no. 2, pp. 190–202, 2016.