IberSPEECH 2018
21-23 November 2018, Barcelona, Spain

Towards an Automatic Evaluation of the Prosody of Individuals with Down Syndrome.

Mario Corrales-Astorgano[1], Pastora Martínez-Castilla[2], David Escudero-Mancebo[1], Lourdes Aguilar[3], César González-Ferreras[1], Valentín Cardeñoso-Payo[1]

[1]Department of Computer Science, University of Valladolid, Spain
[2]Department of Developmental and Educational Psychology, UNED, Spain
[3]Departament of Hispanic Philology, Universitat Autònoma de Barcelona, Spain

mcorrales@infor.uva.es, pastora.martinez@psi.uned.es, descuder@infor.uva.es, lourdes.aguilar@uab.cat, cesargf@infor.uva.es, valen@infor.uva.es

Abstract

Prosodic skills may be powerful to improve the communication of individuals with intellectual and developmental disabilities. Yet, the development of technological resources that consider these skills has received little attention. One reason that explains this gap is the difficulty of including an automatic assessment of prosody that considers the high number of variables and heterogeneity of such individuals. In this work, we propose an approach to predict prosodic quality that will serve as a baseline for future work. A therapist and an expert in prosody judged the prosodic appropriateness of individuals with Down syndrome' speech samples collected with a video game. The judgments of the expert were used to train an automatic classifier that predicts the quality by using acoustic information extracted from the corpus. The best results were obtained with an SVM classifier, with a classification rate of 79.30%. The difficulty of the task is evidenced by the high inter-human rater disagreement, justified by the speakers' heterogeneity and the evaluation conditions. Although only 10% of the oral productions judged as correct by the referees were classified as incorrect by the automatic classifier, a specific analysis with bigger corpora and reference recordings of people with typical development is necessary.

Index Terms: Prosody, Automatic Classification, Down syndrome, Educational Video games

1. Introduction

The collective of individuals with Down syndrome shows a series of cognitive, learning and attentional limitations. All the areas of language are altered, but not in the same degree, as described in [1]. Although lexical acquisition is delayed, morphology and syntax appear to be more affected than vocabulary [2]. Related to pragmatics, individuals with Down syndrome show difficulties when producing and understanding questions and emotions, signaling turn-taking, or keeping topics in conversation, and the study in [3] demonstrated that children with Down syndrome are impaired relative to norms from typically developing children in all areas of pragmatics. At phonological level, speech intelligibility is seriously damaged by the presence of errors on producing some phonemes, the loss of consonants and the simplification of syllables [4].

What concerns to prosody, [5] report disfluencies (stuttering and cluttering) and impairments in the perception, imitation and spontaneous production of prosodic features; authors of [6] have connected some of the speech errors with difficulties in the identification of boundaries between words and sentences. Nevertheless, characterizing prosodic impairments in populations with developmental disorders is a hard task [7]. To fulfill such an aim, prosody assessment procedures appropriate for use with individuals with intellectual and/or developmental disabilities need to be employed. The Profiling Elements of Prosody in Speech-Communication (PEPS-C) test has proved to be successful in this respect [8, 9]. When used with English-speaking children with Down syndrome, lower performance than expected by chronological age is observed in all prosody tasks [10]. After comparisons with typically developing children matched for mental age, impairments are also found for the discrimination and imitation of prosody [10].

There are technological tools focused on language therapy [11, 12]. However, the difficulty of separating the effects of each of the suprasegmental features on communication together with the multiplicity of right possibilities to arrive to the same intonational meaning explains that little attention has been paid to the development of technological resources that specifically consider the learning of prosody in students with special needs, specifically in those with Down syndrome. To advance in the line of developing specific resources to minimize the limitations concerning prosody and pragmatics in individuals with Down syndrome, we have developed an educational video game to train prosody, PRADIA: Mistery in the city [13, 14][1].

Although the video game was designed with the aim of training prosody in individuals with Down syndrome, it became a tool to collect their oral productions and thus to construct a prosodic corpus. These aims are achieved thanks to the fact that the main way of interaction of the player with the game is through the voice. To advance in the game, the player must give an adequate oral response in different communicative circumstances, where prosodic features are the most relevant to achieve a correct pragmatic interpretation. In its current version, the video game needs the constant presence of a person (ideally a therapist) who guides the gamer throughout the adventure and who evaluates the success in the resolution of the production activities. The assistance of the therapist has been proved crucial to motivate individuals with Down syndrome. Even so, it would be desirable to improve their autonomy and to help trainers in their therapies with new functionalities by including a module of automatic assessment of prosodic quality.

If we turn to the field of automatic assessment, the attempts

[1]The activities of Down syndrome speech analysis continue (1/2018-12/2020) in the project funded by the Ministerio de Ciencia, Innovación y Universidades and the European Regional Development Fund FEDER (TIN2017-88858-C2-1-R) and in the project funded by Junta de Castilla y León (VA050G18)

10.21437/IberSPEECH.2018-24

of classifying different speech dimensions is well researched, but focused on specific aspects or reduced populations. Some works focus on speech intelligibility of people with aphasia [15] or speech intelligibility in general [16]. Others try to identify speech disorders in children with cleft lip and palate [17]. In addition, speech emotions and autism spectrum disorders recognition have been investigated [18]. The point is that all these works include a subjective evaluation done by experts as a gold standard to train the classification systems.

In this work, we analyse the difficulties of automatically predicting the quality of the prosody of an oral production and propose a new approach that will serve as a baseline for future work. Recordings of individuals with Down syndrome collected in different sessions of use of the educational video game PRADIA: Mystery in the city were used to obtain information about the relevant features needed to make an automatic classification of the productions. The speech corpus obtained along the time of game was judged by a therapist, who evaluated in real time the quality of the oral productions, and by a prosody expert, who did an off-line evaluation. The difference in the experimental procedure will be used to investigate if an automatic system can only rely on prosodic variables to judge the oral productions of the players (offline evaluation), or whether other features related to the game dynamics should also be incorporated in the system. The judgments of the expert are used to train an automatic classifier that predicts quality by using acoustic information extracted from the audios of the corpus.

In section 2, the experimental procedure is described, which includes the procedure for corpus collection, the processing of speech material and the classification of the samples. The results section shows the effectiveness of the procedure, although, at the same time, the difference in the evaluation procedures highlights the need of carefully defining the selection of features in the process of classification. We end the paper with a discussion about the relevance of the results and the conclusions and future work section.

2. Experimental procedure

2.1. Corpus description

The three subcorpora were recorded using the video game, but the version of the video game and the recording context were different. The complete description of each subcorpus can be seen on Table 1.

To build the subcorpus C1, five young adults with DS (mean age 198 months) were recruited from a local Down syndrome Foundation located in Madrid (Spain). To account for the variability often found in individuals with Down syndrome and get measurements of different developmental variables, all of the participants were administered with the following tests. The Peabody Picture Vocabulary Scale-III [19] was used to assess verbal mental age, the forward digit-span subtest included in the Wechsler Intelligence Scale for Children-IV [20] was employed to evaluate verbal short-term memory and Raven's Coloured Progressive Matrices [21] served as a means to measure nonverbal cognitive level. Descriptive characteristics and scores obtained are shown in Table 2. The full PEPS-C battery in its Spanish version [22] was also administered to participants to have specific measurements of prosody level. Mean percentage of success in perception and production PEPS-C tasks is also presented in Table 2. Once these assessments were completed, participants were administered with the PRADIA video game. Each participant used PRADIA for a total duration of 4 hours, distributed in 4 sessions of 1 hour per week. Participants were supported by a speech and language therapist who knew them in advance and was an expert at working with individuals with Down syndrome. The therapist explained the game, helped participants when needed, and took notes about how each session developed. Importantly, the therapist also assessed participants' speech productions and thus monitored their rhythm of progress within the video game.

Table 1: *Corpus description. Concerning the therapist decision, Cont.R (Continue Right) means that the activity was rightly resolved, Cont (Continue) means that the activity was satisfactorily resolved and Rep (Repeat) means that the activity was faultily resolved. Concerning the expert judgment, Right means that the recording was rightly produced and Wrong means that the recording was wrongly produced.*

		Therapist decision (real time)			**Expert judgment (offline)**		
Speaker	**#Utterances**	**Cont.R**	**Cont.**	**Rep.**	**Right**	**Wrong**	**Corpus**
S01	120	70	33	17	87	33	C1
S02	106	90	16	0	81	25	C1
S03	97	93	3	1	78	19	C1
S04	131	19	51	61	75	56	C1
S05	151	21	54	76	77	74	C1
S06	30	x	x	x	19	11	C2
S07	34	x	x	x	13	21	C2
S08	28	x	x	x	23	5	C2
S09	43	x	x	x	20	23	C2
S10	33	x	x	x	29	4	C2
S11	57	x	x	x	31	26	C3
S12	12	x	x	x	7	5	C3
S13	7	x	x	x	2	5	C3
S14	11	x	x	x	3	8	C3
S15	33	x	x	x	19	14	C3
S16	10	x	x	x	6	4	C3
S17	8	x	x	x	5	3	C3
S18	11	x	x	x	6	5	C3
S19	10	x	x	x	6	4	C3
S20	10	x	x	x	6	4	C3
S21	9	x	x	x	1	8	C3
S22	7	x	x	x	3	4	C3
S23	8	x	x	x	3	5	C3
Total	966	293	157	155	465	302	

C2 subcorpus was also recorded using PRADIA software. These recordings were obtained through the video game within one session of software testing with real users. This test session was done in a school of special education located in Valladolid (Spain). Five adults with Down syndrome, aged 18 to 25, participated in this test. The judgments obtained during this game session were discarded for this work because the speech productions were not evaluated by a therapist. The oral productions were judged in an offline mode by the expert in prosody.

C3 subcorpus was recorded using an older version of PRADIA software, the Magic Stone [23], with less types of production activities. Eighteen young adults with Down syndrome participated in the different game sessions, which focused on how these users interacted with the video game. Five of these eighteen speakers participated as well in the recordings of the C2 subcorpus, so their productions were discarded from C3 subcorpus. As well as in the C2 subcorpus, the judgments decided by the assistant that helped players complete the adventure were not considered in the classifications. Instead, the oral productions were judged in an offline mode by the expert in prosody.

2.2. Corpus evaluation

During the game sessions, a speech therapist sits next to the player and evaluates the production activities in real time. Consequently, in C1 corpus, the therapist adapted her judgments to both the general developmental level of participants and their

Table 2: *Description of the C1 subcorpus. For each speaker, this table shows Chronological age (CA), Verbal mental age (VA), Short-term verbal memory (STVM), and Non-verbal cognitive level (NVCL). Ages are expressed in months. In addition, the mean percentage of success in perception (MPercT) and production (MProdT) PEPS-C tasks are included.*

Speaker	Gender	CA	VA	STVM	NVCL	MPercT	MProdT
S01	f	195	84	94	17	69.79%	48.30%
S02	m	204	99	134	18	76.04%	72.10%
S03	f	178	96	78	20	73.96%	74.65%
S04	m	190	60	below 74	10	60.42%	49.76%
S05	m	223	69	below 74	13	56.25%	54.84%

emotional and motivational level. The video game allows to evaluate the result of the oral activities typing a concrete key on the computer keyboard where the game is installed. If the evaluation is Cont.R (Continue with right result) or Cont (Continue but the oral activity could be better), the video game advances to the next activity. If the evaluation is Rep. (Repeat), the game offers a new attempt in which the player has to repeat the activity. For each activity, there is a predetermined number of attempts: when the attempts finish, the video game goes to the next screen to avoid frustration on the player, even if the activity has not been successfully completed (and the therapist continues judging with Rep.).

On the other hand, an expert in prosody evaluated the three subcorpora of oral productions of 23 speakers with Down syndrome in an offline mode. Due to the difficulty of the task and the different context of the evaluation, the prosody expert used a reduced evaluation system (Right or Wrong production). The judgments were made relying on purely auditive basis, without any acoustic analysis of the sentences, and the focus was on the intonational and prosodic structure. Related to this, factors of intelligibility, quality in pronunciation or adjustment to the expected sentence were not taken into account. Even in the case of speakers with low cognitive level and serious problems of intelligibility, the main criterion was whether they had modeled prosody with certain success, even if the message was not understood. Following the categories of intonational phonology [24] and the learning objectives included in PRADIA [14], criteria concerning intonation, accent and prosodic organization were used to judge if the sentence was Right or Wrong: in short, adjustment to the expected modality; respect for the difference between lexical stress and accent (tonal prominence); and adjustment to the organization in prosodic groups relying mainly in the distinction between function and content words.

2.3. Feature extraction

The openSmile toolkit [25] was used to extract acoustic features from each recording of C1, C2 and C3 subcorpora. The GeMAPS feature set [26] was selected due to the variety of acoustic and prosodic features contained in this set. This set contains frequency related features, energy related features, spectral features and temporal features. The arithmetic mean and the coefficient of variation were calculated on these features. Furthermore, 4 additional temporal features were added: the silence and sounding percentages, silences per second and the mean silences. The complete description of these features can be found in previous research [27]. In this work, only prosodic features (frequency, energy and temporal) have been used because spectral features improve the speaker identification, and classifiers can be adapted to each speaker in the classification process. In total, 34 prosodic features were employed.

2.4. Automatic classification

As explained is section 2.2, the recordings were evaluated by the therapist and the prosody expert. Since the final aim of the module is to decide if the gamer can continue the game or should repeat the activity (without considering degrees of failure), the evaluation of the expert was used to build the classifier. According to this, the output of the different classifiers are Right (R) or Wrong (W), based on the prosody expert scoring. The Weka machine learning toolkit [28] was used and three different classifiers were used to compare their performance: the C4.5 decision tree (DT), the multilayer perceptron (MLP) and the support vector machine (SVM). In addition, the results of using the recordings of the three corpora as well as all combinations of these corpora were compared.

Furthermore, the stratified 10-fold cross-validation technique was used to create the training and testing datasets. We also used feature selection before training the classifiers: the features were selected by measuring the information gain of the training set and discarding the ones in which the information gain equals zero (column Feat. in Table 3).

3. Results

Table 1 and Table 2 show a high difference between speakers related to their developmental level and prosodic skills. S04 and S05 have the lowest scores in *verbal mental age* (60 and 69, respectively), *short-term verbal memory* (below 74 both speakers) and *non-verbal cognitive level* (10 and 13, respectively). In addition, both of them have the lowest mean percentage of success in perception PEPS-C tasks (60.42% and 56.25%, respectively) and lower mean percentage of success in production PEPS-C tasks (49.76% and 54.84%, respectively). These low scores are related with the quality of the productions, with a higher percentage of *W* assignments from the prosody expert (42.75% and 49% respectively) and higher percentage of *Rep.* from the therapist (47% and 50%, respectively).

The classification results highly depend on the corpus and the classifier used (Table 3). SVM classifier works better with all corpora and the worst results are obtained using DT classifier (best case is 79.3% vs 64.94% baseline). The best results are obtained in Case A and D by using any of the three classifiers (UAR 0.83 with SVM classifier). The classification accuracy decreases when the C3 corpus is entered (C, E, F and G cases) as the number of speakers substantially increases. Moreover, when the same features are used to identify speakers instead of the quality of the utterance (column #SR rate of Table 3), scenarios Case C, E and G are the worst ones and scenarios Case A and B are the best. In order to see the influence of the speaker in the classification results, we present results per speaker in Table 4.

We focus on Case D to present results per speakers in Table 4. Only the samples of corpus C1 are analyzed because they were evaluated by the two evaluators. Comparing the R-W judgments of the expert with the classifier predictions, there is a high recall in R-R case for all speakers (S01 83.91%, S02 87.65%, S03 97.44%, S04 94.67%, S05 87.01%). The coincidence in W-W case is lower: while S02 and S05 present a reasonable classification rate (72% and 70.27%, respectively), results for S03 goes down to 26.32%. Concerning this result, we note that most of the utterances judged as wrong by the expert were rated as right by the therapist (100% in cell W-Cont.R for S3). As average, we obtain only 10.05% of false negatives. This will be discussed in the next section as a positive result for

Table 3: *Classification results depending on the corpus and the classifier used. The prosody expert judgments were used to train the classifiers. BL means the performance baseline of each group of samples (number of samples of the most populated class divided by all the samples). DT means Decision trees, SVM means Support vector machines and MLP means Multilayer Perceptron. CR means the classification rate, AUC means the Area Under the Curve and AUR means the Unweighted Average Recall. The number of samples (utt.), the number of speakers (SPK), the number of features (Feat.) and the speaker classification rate using SVM (SR rate) are presented. The output of the different classifiers are Right or Wrong, based on prosody expert scoring.*

			DT			**SVM**			**MLP**						
	Corpora	BL	CR	AUC	UAR	CR	AUC	UAR	CR	AUC	UAR	#Utt.	#Feat.	#SPK	SR rate
Case A	C1	65.79%	69.57%	0.68	0.74	78.49%	0.74	0.83	73.23%	0.7	0.79	605	21	5	69.92%
Case B	C2	61.90%	60.26%	0.58	0.61	72.68%	0.7	0.79	68.49%	0.67	0.73	168	16	5	88.01%
Case C	C3	50.78%	65.76%	0.66	0.66	61.58%	0.62	0.69	63.71%	0.64	0.64	193	7	13	30.05%
Case D	C1+C2	64.94%	70.77%	0.68	0.75	79.3%	0.76	0.83	72.57%	0.7	0.78	773	21	10	64.94%
Case E	C1+C3	62.16%	66.29%	0.65	0.69	72.31%	0.7	0.79	67.17%	0.65	0.74	798	20	18	52.26%
Case F	C2+C3	55.96%	60.94%	0.6	0.64	66.47%	0.66	0.75	64%	0.63	0.69	361	13	18	64.27%
Case G	C1+C2+C3	62.11%	66.88%	0.66	0.71	74.32%	0.71	0.81	69.37%	0.66	0.76	996	20	23	59.21%

real time situations.

Concerning the therapist judgments, *Cont.R* decision could be identified as a *Right* assignment in a high percentage of cases for S01, S02 and S03 speakers (69%, 85% and 95% respectively). They are the participants with higher developmental level, according to Table 2. Among these three participants, the first one -with the lowest inter-judge agreement- showed the lowest prosodic level from the outset. In general, the correspondence between real time decisions and expert judgment is not straightforward, with a high variety in the contingency table. Concerning the therapist *Rep.* decision, it is clear that the highest percentages of agreement are obtained for S04 and S05 speakers (62.5% and 72.97%, respectively), who are the least qualified speakers in Table 2.

Table 4: *Percentage of coincidence between therapist decision, classifier (SVM in case D) 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%	80.20%	68.79%	35.48%
		W	207	41.06%	58.94%	19.80%	31.21%	64.52%

4. Discussion and Conclusions

The study shows some of the variables that contribute to account for the difficulties of conducting an automatic evaluation of prosody in Down syndrome. As shown in Table 2, the chronological age of the participants for whom both the therapist and prosody expert evaluations were available was similar. However, their skills for reasoning, recalling auditory verbal material and understanding vocabulary were clearly different. Phenotype variability is common in Down syndrome [3] and needs to be considered if prosody is to be evaluated. When developmental level is low, the quality of the prosodic productions is also low. As a result, the likelihood of human agreement as to the appropriateness of the output decreases. This shows the difficulties inherent to the task being carried out. Furthermore, even in the cases of higher cognitive level, variability in the linguistic profile can also play a role. Thus, levels of vocabulary are not necessarily paired with those of prosody perception and production. Differences in the evaluation context also explain the variability between the expert and therapist's judgments. While the former only based her decisions on intonational criteria, the latter also took into consideration the progress of the player within 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 our video game, not to evaluate as wrong a right utterance is very important; otherwise, frustration may arise. This is even more important when individuals with Down syndrome are the players since they can be particularly prompted to this feeling [29]. Therefore, one of the main aims of the video game is to engage and motivate the users. For this, the percentage of false positives must be as low as possible. Table 4 shows that only 10% of the samples evaluated as Right by the expert are classified as Wrong by the classifier. It is future work to reduce the rate of false positives in order to obtain the best possible reliable evaluation system.

The differences between the therapist and prosody expert evaluations highlight the importance of evaluation contexts. If the automatic evaluation module aims to be included in a real time video game, aspects different from prosody should be considered, in the line of what the therapist did in her evaluation. The player profile -among other features related to the progress in the game- should also be incorporated in the system. In addition, the evaluation scale can be improved by adding more dimensions to be scored by the experts. Instead of having a global score of the prosody of a recording, the experts could assign a different score to different prosodic dimensions (intonation, pauses, rhythm), with the aim of making a more precise classification.

The high variability of speech of individuals with Down syndrome has been evidenced in experimental results. Further research should compile a bigger and more balanced corpus of the speech of individuals with Down syndrome and record a reference corpus of people with typical development. Nevertheless, inter-speaker variability should be considered as an intrinsic feature of the voices of individuals with Down syndrome so that both the reference of correctness and the particular limitations of the speaker must be taken into account to attain an effective automatic prosodic assessment.

5. References

[1] 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.

[2] P. A. Eadie, M. Fey, J. Douglas, and C. Parsons, "Profiles of grammatical morphology and sentence imitation in children with specific language impairment and down syndrome," *Journal of Speech, Language, and Hearing Research*, vol. 45, no. 4, pp. 720–732, 2002.

[3] E. Smith, K.-A. B. Næss, and C. Jarrold, "Assessing pragmatic communication in children with down syndrome," *Journal of communication disorders*, vol. 68, pp. 10–23, 2017.

[4] G. Laws and D. V. Bishop, "Verbal deficits in down's syndrome and specific language impairment: a comparison," *International Journal of Language & Communication Disorders*, vol. 39, no. 4, pp. 423–451, 2004.

[5] 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.

[6] B. Heselwood, M. Bray, and I. Crookston, "Juncture, rhythm and planning in the speech of an adult with down's syndrome," *Clinical Linguistics & Phonetics*, vol. 9, no. 2, pp. 121–137, 1995.

[7] 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.

[8] P. Martínez-Castilla, M. Sotillo, and R. Campos, "Prosodic abilities of spanish-speaking adolescents and adults with Williams syndrome," *Language and Cognitive Processes*, vol. 26, no. 8, pp. 1055–1082, 2011.

[9] S. Peppé, J. McCann, F. Gibbon, A. O'Hare, and M. Rutherford, "Receptive and expressive prosodic ability in children with high-functioning autism," *Journal of Speech, Language, and Hearing Research*, vol. 50, no. 4, pp. 1015–1028, 2007.

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

[11] 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.

[12] W. R. Rodríguez, O. Saz, and E. Lleida, "A prelingual tool for the education of altered voices," *Speech Communication*, vol. 54, no. 5, pp. 583–600, 2012.

[13] "Pradia," http://www.pradia.net, accessed: 2018-07-18.

[14] L. Aguilar and Gutiérrez-González, "Aprendizaje prosódico en un videojuego educativo dirigido a personas con síndrome de down: definición de objetivos y diseño de actividades," *Revista de Educación Inclusiva*, under revision.

[15] 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.

[16] 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.

[17] A. Maier, F. Hönig, C. Hacker, M. Schuster, and E. Nöth, "Automatic evaluation of characteristic speech disorders in children with cleft lip and palate," in *Ninth Annual Conference of the International Speech Communication Association*, 2008.

[18] H.-y. Lee, T.-y. Hu, H. Jing, Y.-F. Chang, Y. Tsao, Y.-C. Kao, and T.-L. Pao, "Ensemble of machine learning and acoustic segment model techniques for speech emotion and autism spectrum disorders recognition." in *INTERSPEECH*, 2013, pp. 215–219.

[19] L. Dunn, L. Dunn, and D. Arribas, "Test de vocabulario en imágenes peabody," *Madrid: TEA*, 2006.

[20] S. Corral, D. Arribas, P. Santamaría, M. Sueiro, and J. Pereña, "Escala de inteligencia de Wechsler para niños-IV," *Madrid: TEA Ediciones*, 2005.

[21] J. Raven, J. C. Raven *et al.*, *Test de matrices progresivas: manual/Manual for Raven's progessive matrices and vocabulary scalesTest de matrices progresivas*. Paidós,, 1993, no. 159.9. 072.

[22] P. Martínez-Castilla and S. Peppé, "Developing a test of prosodic ability for speakers of iberian spanish," *Speech Communication*, vol. 50, no. 11-12, pp. 900–915, 2008.

[23] 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*, vol. 33, no. 9, pp. 693–712, 2017.

[24] D. R. Ladd, *Intonational phonology*. Cambridge University Press, 2008.

[25] 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.

[26] 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.

[27] 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.

[28] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

[29] J. Grieco, M. Pulsifer, K. Seligsohn, B. Skotko, and A. Schwartz, "Down syndrome: Cognitive and behavioral functioning across the lifespan," in *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, vol. 169, no. 2. Wiley Online Library, 2015, pp. 135–149.