



Incorporation of a module for automatic prediction of oral productions quality in a learning video game

David Escudero-Mancebo, Valentín Cardeñoso-Payo, Mario Corrales-Astorgano, César González Ferreras, Valle Flóres-Lucas¹, Lourdes Aguilar², Yolanda Martín-de-San-Pablo³, Alfonso Rodríguez-de-Rojas⁴

¹Department of Computer Science, University of Valladolid, Spain

²Spanish Philology Department, Universitat Autònoma de Barcelona, Spain

³Fundación Personas, Valladolid, Spain

⁴Asociación Down Valladolid, Spain

descuder@infor.uva.es, valen@infor.uva.es

Abstract

This document presents the research project TIN2017-88858-C2-1-R of the Spanish Government for the incorporation of a module for automatic prediction of oral productions quality focusing on prosody in a learning video game adapted for Down syndrome speakers. It is our goal to present the project in IberSpeech 2020 as it is the main conference gathering specialists on speech technology in Iberian languages. We present the starting point of the project detailing antecedents of the learning game, the collection of human based evaluations during different testing campaigns of the software, the use of the collected data for training automatic assessment components; its integration in the video game for providing autonomous gaming; and the usability tests. As result, the new version of the video game permits autonomous and semi-supervised training thanks to the devised component for evaluating prosodic quality. The project concludes with relevant contributions to state of the art such as an annotated corpus, an open-source learning game and relevant analysis on the atypical prosodic patterns of Down syndrome speakers.¹

Index Terms: Computer Assisted Pronunciation Training, Prosody, Down syndrome speech, Learning video games.

1. Introduction

This project continues the research line opened in 2014 with the project *La piedra mágica*² supported by Resercaixa and continued in 2016 with the project *Pradia*³ supported by BBVA Humanidades Digitales. The first of these projects served to develop a video game for Down syndrome adolescents to train oral communication, in particular prosody and pragmatics [1]. In the second project, the video game routines were enriched and a clinical evaluation of user's performance was tested with formal evaluation.

The video game has the structure of a graphic adventure game, including conversations with characters and navigating through scenarios [1]. It includes three types of activities: comprehension, production and visual cognitive activities. Comprehension activities are focused on lexical-semantic comprehension and the improvement of prosodic perception in specific contexts. In these activities, players have to choose between

different options to continue a conversation with a game character. Production activities are focused on oral production, so the player is encouraged by the game to train his/her speech, keeping in mind such prosodic aspects as intonation, expression of emotions or syllabic emphasis. In these activities, the player is introduced by the game to different conversations with game characters, where the player has to choose between different options to continue the dialogue or to record some sentences related with the dialogue context, depending on the activity. Finally, the visual activities are introduced to add variety to the game and to practice other skills not related directly with language. All activities have a maximum of attempts to avoid frustration in the players, but this maximum depends on the activity.

First we state the goals of the project as they appear in the original application form, next we report on different relevant project management concerns that has to be taken into account in the section of results which contents are aligned with the main publications derived from the project. We end with conclusions and future work.

In this context, in the first usability tests campaigns we detected that the demand of the game was high because it was an attractive resource both for students and for teachers. Nevertheless the original use of the game required the active participation of a therapist who assisted the intellectual disabled user while playing and decided whether he or she must or not repeat the oral production activity. The need of a specialist seated by the student during the training sessions limited the use of the system so that incorporating an automatic module that play the role of the therapist was a need. The training of such system and its incorporation in the learning game without degrading the quality of the training sessions was the challenge of the project.

2. Goals of the project

The video games developed in previous projects require the permanent assistance of a human trainer (a teacher, a therapist, a relative...) to control and monitor the progress of the game users. The goal of the current project is programing a module to increase the software capabilities so that users can play in an autonomous way supervised by an intelligent tutor. This intelligent system will be responsible to decide on the user to repeat or continue with the training activities giving feedback for him/her to be more competent in oral communication and prosodic skills.

Three stages have been defined: 1) corpus collection of audio of Down syndrome (DS) people playing with the video

¹We would like to thank Ministerio de Economía y Competitividad y Fondos FEDER project key: TIN2017-88858-C2-1-R

²<http://prado.uab.es/recercaixa/>

³<http://www.pradia.net>

game; the DS audios will be rated by professional voice therapists; correcting advices of the therapists will be also monitored. 2) computational models of quality of DS turns will be trained from the audio; a knowledge data base will be compiled with the therapist corrective orders. 3) An expert system is using the compiled information to decide about the user activities in real time.

3. Project management

The IPs of the project (first two authors of this paper) have been responsible of the project management. The participation of the different members of the project is reflected in the list of authors of the different publications related with the achievements. Special mention must be done about the role of Lourdes Aguilar who signed a project proposal originally coordinated with the present one. Her proposal was rejected but has given support in what concerns to aspects related to linguistics, prosody and pragmatics. We had the opportunity to hire Mario Corrales whose technical and research initiative has been crucial in the development of the project.

We have collaborated with the research group of Pastora Martínez from the Department of Psychology of UNED Madrid. She is co-author of several relevant publications in the field of Down syndrome and prosody [2, 3]. She has provided us with a paired corpus of typical-DS speech, obtained while applying the PEPS-C test [4].

As stakeholders we counted with Fundación Personas Valladolid and with ASDOVA (Down syndrome Valladolid Association). Therapists and teachers of these institutions (in the list of coauthors) collaborated in the definition of the evaluation template, performed the evaluation and selected the informants. They also actively participated in the usability tests.

The Ethics Committee assistance and consent was required and one of the partners informed about changes in the regulation with respect to image rights of the informants. We specially thanks the support of Ricard Martínez on the compilation of the informed consents. Finally we receive positive authorization PI 20-1639.

The lock-down due to the pandemic COVID-19 interrupted our evaluation sessions in the ASDOVA center. Both the collection of the evaluation template sessions and the usability test campaign had to be postponed. Finally we could perform them during october/november 2020 but we had to ask for a 6 months prorogation so that the project is still alive until June 2021.

4. Achievements

Different parts of this sections are aligned with different articles written during the project development. Some of them are already published, like the ones presented in sections 4.1.1, 4.1.2 and 4.3.2; others are under revision like the ones presented in sections 4.1.3 or 4.3.1; and the one presented in 4.2 is in process.

4.1. Evaluation of prosodic quality of Down syndrome speakers

The works presented in this section are all of them related with the analysis of the prosodic features with different goals. Subsection 4.1.1 presents a work in which the acoustic prosodic features of Down syndrome speakers are compared with the ones of typical development speakers; the whole work can be found in [5]. Subsection 4.1.2 presents an analysis of the quality of the

oral production of Down syndrome speakers from the prosodic acoustic features; the whole work can be found in [6] and in [7]. Finally, 4.1.3 presents an analysis of the prosodic patterns frequently observed in Down syndrome speech when different prosodic functions are performed; further details can be found in [8].

4.1.1. Differences between Down syndrome and typical development speakers

There are many studies that identify important deficits in the voice production of people with Down syndrome. These deficits affect not only the spectral domain, but also the intonation, accent, rhythm and speech rate. The main aim of this work was the identification of the acoustic features that characterize the speech of people with Down syndrome, taking into account the different frequency, energy, temporal and spectral domains. The comparison of the relative weight of these features for the characterization of Down syndrome people's speech was another aim of this study.

The openSmile toolkit with the GeMAPS feature set was used to extract acoustic features from a speech corpus of utterances from typically developing individuals and individuals with Down syndrome. Then, the most discriminant features were identified using statistical tests. Moreover, three binary classifiers were trained using these features.

The best classification rate, using only spectral features, is 87.33%, and using frequency, energy and temporal features, it is 91.83%. Finally, a perception test was performed using recordings created with a prosody transfer algorithm: the prosody of utterances from one group of speakers was transferred to utterances of another group.

The results of this test show the importance of intonation and rhythm in the identification of a voice as non typical. As conclusion, the results obtained point to the training of prosody in order to improve the quality of the speech production of those with Down syndrome.

4.1.2. Analysis of the impact of the speaker in the evaluation

Prosodic skills are useful for improving 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 the heterogeneity of such individuals.

In this work, we analysed how the heterogeneity of people with Down syndrome can affect the automatic assessment of prosodic quality. To do this, 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, with a classification rate of 79.3%.

In addition, the relationship of some prosodic features with the expert assessment of five speakers was also analyzed. We observe the different importance of the prosodic features in the automatic classification of the recordings of each speaker.

This result seems to indicate that this heterogeneity must be taken into account when developing an automatic assessment of the prosodic quality of people with Down syndrome.



Figure 1: *Evaluation session with real users.*

4.1.3. Analysis of the impact of the prosodic function

The speech of people with Down syndrome shows prosodic features which are distinct from the ones observed in the oral productions of typically developing speakers. Although a different prosodic realization does not necessarily imply wrong expression of prosodic functions, atypical expression may hinder communication skills. To ascertain whether this can be the case in individuals with DS is the focus of this work.

We analyzed the acoustic features that better characterize utterances of speakers with Down syndrome when expressing prosodic functions related to emotion, turn-end and phrasal chunking, and compare them with those used by typical development speakers. An oral corpus of speech utterances has been recorded using the PEPS-C prosodic competence evaluation tool.

We used automatic classifiers to prove that the prosodic features that better predict prosodic functions in typical development speakers are less informative in speakers with DS.

Although atypical features are observed in speakers with DS when producing prosodic functions, the intended prosodic function can be identified by listeners and, in most cases, the features correctly discriminate the function with analytical methods. However, a greater difference between the minimal pairs presented in the PEPS-C test is found for typical development speakers in comparison with DS speakers.

The proposed methodological approach provides, on the one hand, the identification of the set of features that distinguish the prosodic productions of Down syndrome and typical development speakers and, on the other, the set of target features for therapy of speakers with DS, from an analysis of the separation of prosodic functions.

4.2. Incorporation of the automatic module in the video game and usability tests

The use of information technologies is broadly extended among the population with intellectual disabilities, also, in lower degree, the use of tools for learning, including learning games. In spite of the use of learning games is widely accepted by the community, due to its great engaging capabilities, there are few works showing its efficiency.

In this work we present the study of the efficiency of the developed learning game by comparing the usability of the system in training sessions with different degrees of supervision of the player by the therapist.

We use the PRADIA software, the learning video game for the training of oral productions (prosody and pragmatics) that

showed high user satisfaction rates in previous works [1] when it was used with the assistance of a therapist. We included a module that provides automatic evaluation of the players oral productions allowing autonomous use of the tool. The use of the game was compared in three scenarios: supervised, autonomous and semi-autonomous (groups of students playing in parallel with the assistance of a teacher).

Different instruments of usability evaluation reveal that, in spite of there are no differences in the degree of engaging, there could be differences in the profiting of the training sessions: lower quality of the recorded audios and more errors in the more autonomous playing modes.

We conclude that Down syndrome people autonomous training is possible (the main goal of the project), implying saving human resources costs, but the performance is highly dependent on the feedback provided. In spite of the degree of engagement of the autonomous version is high, the quality and quantity of feedback is not comparable with the one provided by the therapists resulting on lower performance in the training sessions.

4.3. Resources

The works presented in this section have to do with the resources generated during the project. They are freely available for the research community and they consist of an annotated corpus of Down syndrome utterances and the video game itself. Subsection 4.3.1 presents the corpus (further details will be found in [9]); and subsection 4.3.2 the video game with details in [10] and [11].

4.3.1. The corpus of Down syndrome recordings

Oral productions of speakers with Down syndrome exhibit special characteristics that have been the target of study for decades. In spite of this attention, the availability of rich resources for its analysis is still scarce. In this project, we present the definition and compiling procedure of a corpus of semi-controlled oral productions of speakers with Down syndrome that aims to allow the analysis of how speakers with Down syndrome produce functional and linguistic aspects of speech.

The corpus (named PRAUTOCAL) has been recorded while speakers with Down syndrome use a video game for training oral competences. Utterances are related to well defined communicative tasks recorded by both speakers with Down syndrome and typically developing speakers. We present the procedure for human experts to evaluate the recordings and the transcription criteria followed for enriching the utterances of the corpus.

Although the activities of the video game mainly focus on prosody and pragmatics, we show that PRAUTOCAL permits the analysis of the clear contrast in voice and speech between individuals with Down syndrome and typically developing speakers, taking into account the high heterogeneity of the speech problems characteristic of the syndrome.

This material allows the analysis of the speech problems characteristic of the syndrome with applications to the generation of knowledge of the particular problem of these speakers that could be used in future works for therapists to prepare specific training or enriching diagnosis regarding possible speech and language disorders.

4.3.2. Dissemination of the video game

With the growth in popularity of video games in our society many teachers have worked to incorporate gaming into their classroom. It is generally agreed that by adding something fun to the learning process students become more engaged and, consequently, retain more knowledge. However, although the characteristics of video games facilitate the dynamics of the educational process it is necessary to plan a pedagogical project that includes delimitation of learning goals and profile of the addressees, the conditions of application of the educational project, and the methodologies of evaluation of the learning progress.

This is how we can make a real difference between gamification and video game based learning. The paper addresses the design of an educational resource for special education needs students that aims to help teach communicative skills related to prosody. The technological choices made to support the pedagogic issues that underlie the educational product, the strategies to convert learning content into playful material, and the methodology to obtain measures of its playability and effectiveness are described.

The results of the motivation test certified that the video game is useful in encouraging the users to exercise their voice and the indicators of the degree of achievement of the learning goals serve to identify the most affected prosodic skills.

5. Conclusions and future work

We conclude that the goal of the the project has been fulfilled with the definition of a new module of the video game that permit autonomous gaming to Down syndrome players. The module for the evaluation of quality of learners oral productions has shown to be effective deciding whether players can continue playing or have to repeat the activities. Additionally, it has been done without degrading the engagement of users in the video game.

This project contribute to the building of knowledge about Down syndrome speech and voice, with the identification of acoustic prosodic features that better characterize it; we have identified atypical prosodic patterns used by this type of speakers and we have devised strategies to face up this peculiarities in the challenging frame of automatic assessment of prosody.

We are still working on the improvement of module for quality assessment. In particular, until the end of the project (June 2021) we are using the data collected in the usability tests to test and train a classifier that is specialized in the particular activity that is presented by the video game. The learning video game presents different activities for training different prosodic and pragmatic functions. Experimental results [12] show that they must be taken into account for being efficient on the evaluation of prosodic quality.

It has been applied a new research project to the Spanish Ministry for continuing this research line. In particular, we want to deepen into user adaptation concerns. Experimental results have shown that evaluation of prosodic quality highly depends on the particular speaker that is under analysis as DS speakers present a broad spectrum of not only speech deficits but also short memory limitations and intellectual disability. Our proposal is using the game records of the user for categorizing him/her and adapting the game experience in function of the game profile. Important benefits in terms of usability are expected to be obtained with this promising approach.

6. Acknowledgments

Special thanks to Jesús Gómez from CEIP Urueña for the collaboration during the evaluation template definition and during the inter-rater consistency tests.

7. References

- [1] 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.
- [2] 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.
- [3] S. J. Peppé, P. Martínez-Castilla, M. Coene, I. Hesling, I. Moen, and F. Gibbon, "Assessing prosodic skills in five european languages: Cross-linguistic differences in typical and atypical populations," *International journal of speech-language pathology*, vol. 12, no. 1, pp. 1–7, 2010.
- [4] S. Peppé and J. McCann, "Assessing intonation and prosody in children with atypical language development: the peps-c test and the revised version," *Clinical Linguistics & Phonetics*, vol. 17, no. 4-5, pp. 345–354, 2003.
- [5] 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.
- [6] 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.
- [7] M. Corrales-Astorgano, P. Martínez-Castilla, D. Escudero-Mancebo, L. Aguilar, C. González-Ferreras, and V. Cardeñoso-Payo, "Towards an automatic evaluation of the prosody of people with Down syndrome," in *Proc. IberSPEECH 2018*, 2018, pp. 112–116. [Online]. Available: <http://dx.doi.org/10.21437/IberSPEECH.2018-24>
- [8] M. Corrales-Astorgano, D. Escudero-Mancebo, C. González-Ferreras, V. C. Payo, and P. Martínez-Castilla, "Analysis of atypical prosodic patterns in the speech of people with down syndrome," *Biomedical Signal Processing and Control*, p. under evaluation, 2021.
- [9] D. Escudero-Mancebo, M. Corrales-Astorgano, V. Cardeñoso-Payo, L. Aguilar, C. González-Ferreras, P. Martínez-Castilla, and V. Flores-Lucas, "Prautocal corpus:a corpus for the study of Down syndrome prosodic aspects," *Language Resources and Evaluation*, p. under evaluation, 2021.
- [10] L. Aguilar, "Learning prosody in a video game-based learning approach," *Multimodal Technologies and Interaction*, vol. 3, no. 3, p. 51, 2019.
- [11] F. Adell, L. Aguilar, M. Corrales-Astorgano, and D. Escudero-Mancebo, "Proceso de innovación educativa en educación especial: Enseñanza de la prosodia con fines comunicativos con el apoyo de un videojuego educativo," *Humanidades Digitales, Retos, Recursos y Nuevas Propuestas*, pp. 277–293, 2018.
- [12] D. Escudero-Mancebo, M. Corrales-Astorgano, C. González-Ferreras, and V. Cardenoso-Payo, "Prosodic feature selection for automatic quality assessment of oral productions in people with Down syndrome," *Iberspeech 2020*, p. under revision, 2021.