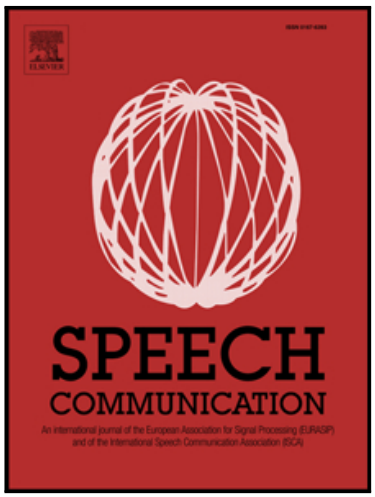

Acoustic characterization and perceptual analysis of the relative importance of prosody in speech of people with Down syndrome

Mario Corrales-Astorgano, David Escudero-Mancebo, César González-Ferreras

PII: S0167-6393(17)30466-1
DOI: 10.1016/j.specom.2018.03.006
Reference: SPECOM 2546

To appear in: *Speech Communication*

Received date: 18 December 2017
Revised date: 21 February 2018
Accepted date: 13 March 2018

Please cite this article as: Mario Corrales-Astorgano, David Escudero-Mancebo, César González-Ferreras, Acoustic characterization and perceptual analysis of the relative importance of prosody in speech of people with Down syndrome, *Speech Communication* (2018), doi: 10.1016/j.specom.2018.03.006

Acoustic characterization and perceptual analysis of the relative importance of prosody in speech of people with Down syndrome

Mario Corrales-Astorgano[a,*], David Escudero-Mancebo[a], César González-Ferreras[a]

[a] *Departamento de Informática, Universidad de Valladolid, Valladolid, Spain*

Abstract

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 is 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 is 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 has been 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.

Keywords: Speech characterization, Prosody, Down syndrome, Intellectual disabilities, Automatic classification, Perceptual test

1. Introduction

Individuals with Down syndrome (DS) have problems in their language development that make their social relationships and their developmental ability more problematic (Cleland et al., 2010; Martin et al., 2009; Chapman, 1997). Many DS individuals have some physiological peculiarities that affect their voice production, such as a smaller vocal tract with respect to the tongue size or soft palatal shape, among others Guimaraes et al. (2008). Muscular hypotonia also affects their capabilities for performing a correct articulation, degrading the quality of the spectral characteristics of sounds (Markaki and Stylianou, 2011). In addition, hearing loss during childhood (Shott et al., 2001) and fluency deficits (Devenny and Silverman, 1990) influence the frequency, energy and temporal domains of the voice signal.

Although problems derived from physiological peculiarities are permanent (only surgery (Leshin, 2000) or prostheses (Bhagyalakshmi et al., 2007) could improve them), intonation and fluency deficits can be improved by speech therapy and training. There are tools available for this goal (González-Ferreras et al., 2017) based on perception and production activities to be performed with the assistance of therapists who help patients to properly manage their breathing and intonation patterns. Although there is general consensus about the importance of improving prosody by training (see (Kent and Vorperian, 2013) for a complete state of art revision), there are very few works that provide empirical evidence of the importance of the prosody related features (those belonging to fundamental frequency, energy and duration domains) with respect to other acoustic features belonging to the spectral domain.

The use of the video game described by González-Ferreras et al. (2017) has allowed the formation of a speech corpus, which has been used in this work to analyze and characterize the speech of people with Down syndrome. This corpus, described in section 3.1, contains recordings of people with Down syndrome and typically developing people. Both groups recorded the same sentences, so statistical and perceptual tests have been used to compare the acoustic features of the two groups of speakers, so that the most relevant differences could be identified.

This work aims to find the best acoustic features to characterize the speech of people with Down syndrome. To do this, features of frequency, energy, temporal and spectral domains have been extracted from the recordings

*Corresponding author

Email addresses: mcorrales@infor.uva.es (Mario Corrales-Astorgano), descuder@infor.uva.es (David Escudero-Mancebo), cesargf@infor.uva.es (César González-Ferreras)