

A first approach to Acoustic Characterization of Costa Rican Children's Speech


Un primer acercamiento a la caracterización acústica del habla de niños costarricenses

Marvin Coto-Jiménez¹, Maribel Morales-Rodríguez²,
Daniel Vargas-Díaz³


Coto-Jiménez, M; Morales-Rodríguez, M; Vargas-Díaz, D.
A first approach to Acoustic Characterization of Costa Rican Children's Speech. *Tecnología en Marcha*. Edición especial 2020. 6th Latin America High Performance Computing Conference (CARLA). Pág 80-84.

 <https://doi.org/10.18845/tm.v33i5.5080>


1 PRIS-Lab: Pattern Recognition and Intelligent Systems Laboratory, Department of Electrical Engineering, School of Engineering, Universidad de Costa Rica (UCR). E-mail: marvin.coto@ucr.ac.cr.

 <https://orcid.org/0000-0002-6833-9938>

2 Department of Orientation and Special Education, Universidad de Costa Rica (UCR). E-mail: maribel.moralesrodriguez@ucr.ac.cr.

 <https://orcid.org/0000-0002-3426-5192>

3 PRIS-Lab: Pattern Recognition and Intelligent Systems Laboratory, Department of Electrical Engineering, School of Engineering, Universidad de Costa Rica (UCR). E-mail: daniel.vargasdiaz@ucr.ac.cr.

 <https://orcid.org/0000-0002-9015-2328>



Keywords

Formants; Signal processing; speech technologies.

Abstract

As human interaction with computers becomes more pervasive, the value of developing automatic speech recognition, text-to-speech synthesis, and related speech technologies become more important for people of all ages, accents, and conditions.

One of the groups that represent bigger challenges is children, due to the difficulties in recording enough speech, and the lack of characterization of their speech, which is particular of every language and accent. This paper presents the first approach to acoustic analyses of Costa Rican children aged from six to twelve years. These analyses aimed to achieve a better understanding of the characteristics of speech produced by this group, in terms of providing future development and enhancement of automatic speech recognizers and speaker identification systems.

For this purpose, we record the speech consisting of isolated words of three children, and compare the results with three adults, in terms of the vowel's formants. The formants give information about the vocal track of the speaker, and it is an important method to provide the first analysis of these signals. Results show noticeable differences between the children and adults and may provide useful information about future trends to adapt and develop the current speech technologies for this population.

Palabras clave

Formantes; procesamiento de señales; tecnologías del habla.

Resumen

A medida que la interacción de las personas con las computadoras se hace más extendida, se vuelve más importante el desarrollo de tecnologías para el reconocimiento automático de la voz, la síntesis de voz, así como otras tecnologías relacionadas, considerando a personas de todas las edades, acentos y condiciones. Uno de los grupos humanos que representan desafíos más grandes es el de los niños, debido a las dificultades para grabar suficientes recursos de habla, y la falta de caracterización de su forma de hablar, la cual es particular de cada idioma y acento. Este artículo presenta una primera aproximación para el análisis acústico de niños costarricenses de entre seis y doce años. Estos análisis tienen como objetivo lograr una mejor comprensión de las características del habla producida por este grupo en particular, en términos de propiciar el desarrollo y mejora de los reconocedores automáticos del habla y los sistemas de identificación del hablante.

Para este propósito realizamos grabaciones de habla de tres niños, las cuales consistieron en palabras aisladas, y comparamos los resultados con tres adultos en términos de los formantes de las vocales. Los formantes proporcionan información sobre el tracto vocal del hablante, y es un parámetro importante para establecer un primer análisis de estas señales. Los resultados muestran diferencias notables entre los niños y los adultos y pueden brindar información útil para futuros estudios en términos de adaptar y desarrollar las tecnologías del habla para esta población.

Introduction

Human interaction with computers and technological devices of all kind has become more extensive in recent years. Being the speech the main form of human communication, the value of speech technologies, such as Automatic Speech Recognition (ASR), Text-to-speech synthesis and related technologies has increased.

This remarkable importance has still many challenges in building robust, and more natural systems for all people, including the elderly, children and people with disabilities. In the case of children, using speech technology is a field underdeveloped in certain contexts, as in the case of Costa Rica.

Several previous analyses have been focused on acoustic analysis of certain sounds, being the most frequent the vowels. Moreover, since this has been performed not only to age-dependent characteristics but also for specific accents [1] or conditions, such as Parkinson's disease [2].

For example, several references have reported the differences between acoustic and linguistic characteristics of children's speech [3-4]. In the English language, children's speech is characterized by higher pitch, and formants occur at higher frequencies [5]. This has an impact if exists on automatic speech recognition or analysis in the presence of perturbations or degradation of the signal, such as bandwidth reduction. Also, in the English language, children below the age of 10 exhibits a wider range of vowel durations relative to older children and adults, and wider variability in formant locations.

If automatic speech recognition systems are trained using acoustic models from adult speech and tested against speech from children, show performance degradation with decreasing age. On average, the word error rates are two to five times worse for children speech than for adult speech [6].

In this work, we perform a first approach to acoustic characterization of Costa Rican children speech, to achieve a better understanding of this particular age group.

Materials and methods

Like other Latin American variants of the Spanish language, the Costa Rican Spanish has five main categories of vowels [7]: <a> (open), <e> (open), <i> (closed), <o> (open), <u> (closed). The open and close categories refer to the general classification of vowels based on whether the sound is produced with the tongue far from the roof of the mouth (open) or with the tongue touching the roof of the mouth (closed).

To provide the inputs for the children voice database, several recording sessions were made during this study. The first session was held on January 2019 and had the participation of 3 children between 6 and 12 years old. The gender and ages of the participants are described in tables 1 and 2.

Table 1. Age of participants (children)

Gender	Age (years)
Male	6
Female	8
Female	12

Table 2. Age of participants (adults)

Gender	Age (years)
Male	18
Female	20
Female	23

Results

Previous references have related the changes in the formant of vowels with the characteristics and form of the vocal tract [8]. The vowel space is used to compare the range and variation in the position of the two first formants.

Figure 1 shows the mean position of the first two formants for two female children and two female adults of the dataset, using the same scale on both axes. The children have variable patterns in the polygon of formants. Whereas the adults demonstrate that they have a wider range regarding the formants and form a more homogeneous shape in this polygon.

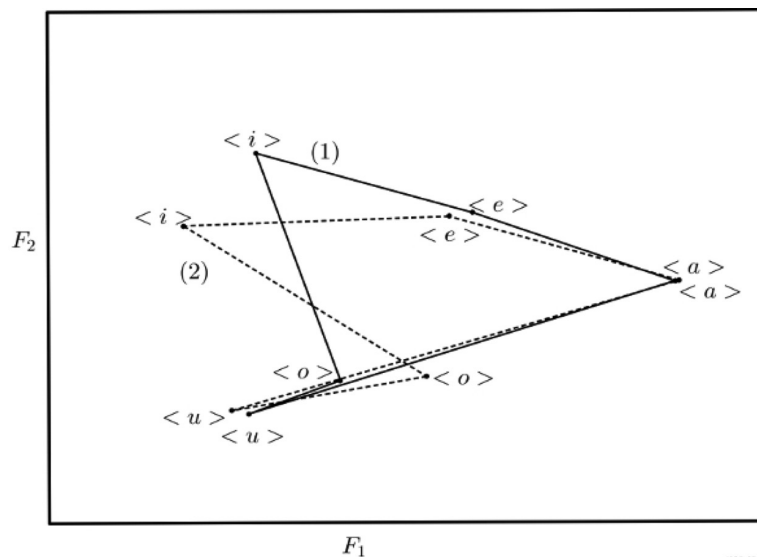


Figure 1: Changes in F1-F2 vowel space as a function of age (Female children (2) and adult speakers (1))

Table 3 shows the mean value of the formants for the <u> vowel. It is noticeable how different the values are between children and adults, even more for the second formant.

Conclusions

In this paper, the first approach to acoustic analyses of Costa Rican children's speech was presented. These analyses were focused on formants. The first group of children 6-12 years was recorded using isolated words, and the words were replied in a group of adults, for comparison purposes.

The formant polygons show different patterns between adults and children. These results are valuable for the knowledge of the speaker characteristics for this population and allow future research in areas such as gender/age voice recognition based on acoustic parameters such as formants. Additionally, the characterization of speech for this age and geographical region may allow the implementation of algorithms to improve the performance of automatic speech recognizers or other speech technologies for this variant of Spanish.

Table 3. Mean value of first and second formants for the vowel <u>. All the values in Hertz.

Speaker	First formant	Second formant
Child (Male, 6)	459,58	1268,76
Child (Female, 8)	441,83	1254,75
Child (Female, 12)	432,94	1192,64
Mean values (children)	444,78	1238,72
Adult (Male, 18)	362,94	1152,33
Adult (Female, 20)	409,20	1103,10
Adult (Female, 23)	444,24	1054,30
Mean values (adults)	405,46	1103,24

Acknowledgments

This work was supported by the University of Costa Rica (UCR), Project No. 322-B9-105, and Project No. ED-3416.

References

- [1] T. Leinonen, *An acoustic analysis of vowel pronunciation in Swedish dialects*. Groningen: [Rijksuniv.], 2010.
- [2] S. Skodda, W. Visser and U. Schlegel, "Vowel Articulation in Parkinson's Disease", *Journal of Voice*, vol. 25, no. 4, pp. 467-472, 2011. Available: 10.1016/j.jvoice.2010.01.009.
- [3] J. McKechnie, B. Ahmed, R. Gutierrez-Osuna, P. Monroe, P. McCabe and K. Ballard, "Automated speech analysis tools for children's speech production: A systematic literature review", *International Journal of Speech-Language Pathology*, vol. 20, no. 6, pp. 583-598, 2018. Available: 10.1080/17549507.2018.1477991.
- [4] L. Perry, M. Perlman, B. Winter, D. Massaro and G. Lupyan, "Iconicity in the speech of children and adults", *Developmental Science*, vol. 21, no. 3, p. e12572, 2017. Available: 10.1111/desc.12572.
- [5] S. Safavi, M. Najafian, A. Hanani, M. Russell, P. Jancovic, P., M. Carey, "Speaker recognition for children's speech". *arXiv preprint*. Available: arXiv:1609.07498.
- [6] A. Potamianos, S. Narayanan, "Robust recognition of children's speech", *IEEE Transactions on speech and audio processing*, vol. 11, no. 6, pp. 603-616.
- [7] F. Martínez-Licona, J. Goddard-Close, A. Martínez-Licona and M. Coto-Jiménez, "Models and Analysis of Vocal Emissions for Biomedical Applications", 2013, pp. 235-238.
- [8] J. Eichhorn, R. Kent, D. Austin and H. Vorperian, "Effects of Aging on Vocal Fundamental Frequency and Vowel Formants in Men and Women", *Journal of Voice*, vol. 32, no. 5, pp. 644.e1-644.e9, 2018. Available: 10.1016/j.jvoice.2017.08.003.