

LINGÜÍSTICA



Revista de Filología y Lingüística de la Universidad de Costa Rica

Publicación Semestral, ISSN: 0377-628X / EISSN: 2215-2628

Volumen 46 - 2

Octubre 2020 - Marzo 2021

**ANÁLISIS ACÚSTICO DE VOCALES DE
NIÑOS COSTARRICENSES**

*Marvin Coto Jiménez
Maribel Morales Rodríguez*



Doi: <https://doi.org/10.15517/rfl.v46i2.42445>

URL: <https://revistas.ucr.ac.cr/index.php/filyling/index>

ANÁLISIS ACÚSTICO DE VOCALES DE NIÑOS COSTARRICENSES

ACOUSTIC VOWEL ANALYSIS OF COSTA RICAN CHILDREN

Marvin Coto Jiménez
Maribel Morales Rodríguez

RESUMEN

En este artículo se presenta un análisis acústico de vocales pronunciadas por niños costarricenses con edades entre los 6 y los 12 años. Estos análisis tienen como objetivo lograr una mejor comprensión de las características del habla producidas por este grupo en particular y, de esta manera, brindar información importante para la mejora de los reconocedores automáticos del habla y la identificación del hablante. Para este propósito, se construyó una base de datos con palabras aisladas que contemplan los fonemas, de acuerdo con varias etapas del desarrollo del lenguaje. Adicionalmente, se desarrolló una estrategia para la interacción con los niños. Los resultados confirman y complementan los resultados reportados en la literatura, pues muestran que existe una mayor variabilidad en parámetros como los formantes, la duración y el tono de las vocales emitidas por los niños, en comparación con los adultos. Los resultados permiten caracterizar cuantitativamente estos sonidos del habla, y establecer futuras investigaciones en áreas como el reconocimiento de habla y su dependencia del género y la edad dentro de esta comunidad.

Palabras clave: análisis del habla; niños; tonos; vocales; fonemas.

ABSTRACT

This paper presents an acoustic analysis of vowels pronounced by a group of Costa Rican children, aged between six and twelve years. These analyses aim to achieve a better understanding of the speech characteristics produced by this particular group and thus provide important information for the improvement of automatic speech recognizers and speaker identification. For this purpose, a database consisting of isolated words was built, which contains the phonemes according to various stages of language development in children. Additionally, a strategy for interaction with them was developed. The results confirm and complement the results reported in the literature, showing that a greater variability of speech parameters such as formants, tone, and duration are present in children in comparison to adults. The results allow the characterization of these speech sounds and the possibility of establishing future research in areas such as speech recognition and its dependence on gender and age in this community.

Keywords: speech analysis; children; pitch; vowels; phonemes.

Dr. Marvin Coto Jiménez. Profesor catedrático. Escuela de Ingeniería Eléctrica, Universidad de Costa Rica. Costa Rica.
Correo electrónico: marvin.coto@ucr.ac.cr

Mag. Maribel Morales Rodríguez. Profesora instructora. Escuela de Orientación y Educación Especial, Universidad de Costa Rica. Costa Rica.

Correo electrónico: maribel.moralesrodriguez@ucr.ac.cr

Recepción: 17- 02- 20

Aceptación: 20- 04- 20

1. Introducción

La interacción humana con computadoras y dispositivos tecnológicos de todo tipo se ha vuelto más extensiva durante los últimos años, en casi todos los países del mundo. Dado que el habla es la forma de comunicación más natural entre las personas, el valor de las tecnologías del habla, entre las que se encuentra el reconocimiento automático del habla (ASR) y la síntesis de voz, ha aumentado.

Esta notable importancia todavía tiene muchos desafíos en la construcción de aplicaciones y sistemas robustos y más naturales para todas las personas; esto incluye, por ejemplo, la elaboración de sistemas de reconocimiento y generación de voces artificiales de niños, niñas y adolescentes, y para personas con discapacidad. Para el caso de las personas menores de edad, el empleo de estas tecnologías podría ser beneficioso en varias áreas de aplicación, como lo son la educación, la seguridad y la protección infantil (Safavi et al., 2012).

Un ejemplo de esto se da en las plataformas de redes sociales que utilizan el reconocimiento de voz. En estas, un sistema puede identificar a un niño a partir de su voz y confirmar la identidad de la persona con la que se comunica, por razones de seguridad. Por su parte, en el área de la educación, un tutor virtual interactivo podría identificar a cada niño de una clase y continuar automáticamente una lección anterior, así como adaptar su contenido a condiciones o situaciones específicas.

Además, el reconocimiento y la síntesis del habla pueden ayudar al desarrollo del habla y el lenguaje en niños pequeños. Por lo tanto, podría apoyar a niños y niñas a mejorar su capacidad de comunicación (Li y Russell, 2002). A pesar de estas ventajas y aplicaciones potenciales, el uso de la tecnología del habla con niños es un campo no desarrollado en ciertos contextos, como sucede en el caso de Costa Rica.

Uno de los primeros pasos para generar sistemas de reconocimiento y síntesis de habla que estén dirigidos a determinados grupos etarios o regionales es el conocimiento de las características particulares de los elementos que conforman la comunicación oral. Estudios anteriores realizados en países como Suecia (Leinonen, 2010) y Estados Unidos (Gerosa, Giuliani y Brugnara, 2007) se han centrado en el análisis acústico de ciertos sonidos, principalmente de las vocales. Estos análisis se han realizado para estudiar aspectos como las características de los acentos específicos o la afectación en el habla ante determinadas condiciones, tales como la enfermedad de Parkinson (Skodda, Visser y Schlegel, 2011) y pólipos de cuerdas vocales (Petrović-Lazić, Babac, Vuković, Kosanović y Ivanković, 2011).

1.1 Objetivo de la investigación

El presente trabajo tiene como objetivo principal describir las características acústicas relacionadas con los formantes, la duración y el tono en una primera muestra tomada de cinco niñas y niños costarricenses, para lograr una mejor comprensión del habla en este grupo de edad en particular y su potencial impacto en áreas de tecnologías del habla, como el reconocimiento de voz. Para establecer una comparación, se incluye un grupo de nueve adultos costarricenses en los cuales se realizan los mismos análisis, de manera que se puedan cuantificar diferencias como la variabilidad que pueda caracterizar la pronunciación de las vocales en los niños.

2. El habla infantil y algunas de sus características

Diversos estudios han informado sobre las diferencias entre las características acústicas y lingüísticas del habla infantil en relación con la de adultos. Por ejemplo, en inglés, el habla de los niños se caracteriza por un tono más alto y por formantes que ocurren en frecuencias más altas (Safavi et al., 2012). Específicamente, los niños menores de 10 años exhiben un rango más amplio de duración de las vocales en relación con los niños mayores y los adultos. También se produce una variabilidad más amplia en las ubicaciones de formantes. La importancia del desarrollo del habla de los niños y niñas se ha observado en ideas como “Tutores de alfabetización interactiva”, donde la necesidad de aumentar las tasas de reconocimiento del habla ha sido señalada claramente (Hagen, Pellom y Cole, 2003).

En el estudio de Potamianos y Narayanan (2003), realizado también en idioma inglés, se determinó que, si los sistemas automáticos de reconocimiento de voz se entrenan utilizando modelos acústicos de adultos y se prueban contra el habla de niños, muestran una degradación en su rendimiento con la disminución de la edad. Esto significa que en otros idiomas se han determinado los problemas que representan para los sistemas de comunicación humano-máquina el hecho de que el usuario sea un niño, en términos de un rendimiento menor, en comparación con los adultos.

Dada la importancia de medir las diferencias presentadas entre el habla de infantes con la de adultos, por su impacto en el reconocimiento automático del habla y demás tecnologías, otros estudios han investigado la duración de los fonemas en consonantes y vocales, con el fin de establecer una correlación con la edad (Gerosa et al., 2007) y con el dominio del habla (Li y Russell, 2002).

La mayoría de la investigación previa sobre el análisis del habla infantil y su impacto en las tecnologías del habla se ha realizado para niños que tienen el inglés como primer idioma. En el caso del castellano se han presentado pocos estudios, incluso después de considerar a los hablantes nativos de español y su impacto en el desarrollo del idioma inglés (You, Alwan, Kazemzadeh y Narayanan, 2005).

En el caso de adultos e idioma castellano sí se han realizado algunas investigaciones como la de Martínez-Licon, Goddard-Close y Coto-Jiménez (2013) y la de Coto-Jiménez, Goddard-Close y Martínez-Licon (2014a), donde se rescatan las características acústicas de aspectos como la duración de las vocales, las frecuencias fundamentales y los dos primeros formantes, los cuales son definidos como los picos destacados en el espectro de los sonidos. También se han explorado características similares para evaluar el habla artificial producida en esta lengua (Coto-Jiménez, Martínez-Licon y Goddard-Close, 2014b). Más recientemente, un estudio cuantificó la diferencia entre el castellano costarricense y el europeo, con relación a la velocidad del habla (Schwab, 2015).

Es bien sabido que las características de una lengua varían de acuerdo con los múltiples acentos que se generan a lo largo del tiempo. En todas las referencias anteriores no se han abordado la descripción ni el análisis del habla en niños latinoamericanos o, en particular, en el acento costarricense. Este hecho nos motiva a desarrollar las metodologías y el análisis acústico del habla, específicamente de las vocales, en este rango de edad y país, para contribuir al conocimiento de estas y al impacto de sus variaciones en las tecnologías del habla.

El estudio realizado en el presente trabajo se puede considerar similar al de Marín Gálvez (1995) y Soto Barba (2007) en cuanto a las características analizadas y el uso de las cinco vocales, pero en una población previamente no estudiada. El tamaño de la muestra también hace comparable el estudio al de Soto Barba (2007), ya que se contó con la participación de cinco menores de edad entre los 6 y 12 años, tanto varones como mujeres, como se mencionó anteriormente.

3. Materiales y métodos

3.1 La obtención del corpus

Con el fin de obtener información sobre el análisis acústico de vocales en español, se realizó una revisión bibliográfica que nutriera con sus hallazgos el presente trabajo. Además, se contó con una metodología mixta, ya que para comprender las características acústicas de los fonemas de las vocales en español producidos por personas menores de edad, se requiere de un escenario naturalista, con una muestra pequeña y cualificada, como lo dicta el enfoque cualitativo, y desde una realidad cuantitativa que dicte una fiabilidad de datos sólidos y replicables.

Como indican Hernández, Fernández y Baptista (2010), “los métodos mixtos representan un conjunto de procesos sistemáticos, empíricos y críticos de investigación e implican la recolección y análisis de datos cuantitativos y cualitativos así como su integración y discusión conjunta” (p. 546).

Para este estudio, se consideró como criterio de selección intencional el rango de edad de los participantes, así como el hecho de contar tanto con niños como con niñas para poder evidenciar las características de la producción acústica del habla en la emisión de vocales, y tener un banco de datos para en posteriores análisis evidenciar similitudes y diferencias en relación al género y la producción oral de los fonemas de las vocales en español.

En cuanto a los datos obtenidos para el análisis de las emisiones acústicas de las vocales, se les dio un tratamiento desde el enfoque cualitativo, en el cual, como lo sugiere Delgado (2014), se hizo uso de lenguaje matemático al asignarle valores numéricos que permiten simbolizar el fenómeno de estudio y la comparación y relación de variables de las emisiones acústicas de las personas participantes, tanto menores de edad como jóvenes adultos estudiantes de las carreras de ingeniería eléctrica y educación especial.

3.2 Vocales analizadas

De forma similar a otras variantes latinoamericanas del idioma español, el español costarricense tiene cinco categorías principales de vocales (Calvo Shadid, 2008): /a/ (abierta), /e/ (abierta), /i/ (cerrada), /o/ (abierta), /u/ (cerrada). Existe un análisis que contempla aún más variantes que surgen de diptongos o hiatos. Para los fines de este trabajo, de carácter exploratorio, se considerarán solamente las cinco categorías de vocales mencionadas.

El estudio de las cinco vocales en su categorización fundamental también ha sido abordado en estudios como el de Marín Gálvez (1995). En este, se determinan algunas condiciones particulares como el acento y la estructura silábica para analizar las diferencias en la duración de cada una en adultos españoles. Como indica el autor “[n]o hemos incluido en el análisis la totalidad de las variables que parecen influir en la duración vocálica en otros

estudios, ya que los resultados aquí presentados constituyen una primera aproximación al tema” (Marín Gálvez, 1995, p. 214). Salamanca Gutiérrez y Valverde San Martín (2009), por su parte, han analizado las cinco vocales en términos de su percepción subjetiva, para determinar el estrato social del hablante en una muestra de adultos chilenos.

En cuanto a formantes, las diferencias han sido establecidas no solamente para distinguir entre las cinco vocales, sino también para utilizarlas como una forma de distinguir variantes del acento chileno de zonas urbanas y rurales (Soto Barba, 2007). La caracterización acústica general de las vocales ha sido presentada también para una región de Chile (Díaz, Cisternas y López, 2015).

En variantes de español mexicano, las vocales también han sido estudiadas recientemente en términos del debilitamiento de las vocales, de acuerdo con su contexto lingüístico y las particularidades de una región específica del país (Serrano, 2006). Estos estudios muestran la relevancia del análisis acústico de las vocales como una forma de caracterizar el habla de determinados acentos particulares y variantes del castellano. Dado que la lengua evoluciona continuamente, es importante la constante actualización de estos estudios.

3.3 Participantes

Para proporcionar las entradas para la base de datos de voz de los participantes menores de edad, se realizaron varias sesiones de grabación durante este estudio. La primera sesión se efectuó en enero de 2019 y contó con la participación de dos niños y tres niñas de entre 6 y 12 años. El género, edad, notación y zona de procedencia de los participantes menores se describe en la Tabla 1. Por su parte, el detalle correspondiente a los participantes adultos que se utilizó para la comparación se muestra en la Tabla 2. En estas tablas, se ha codificado a cada participante en la columna Notación con cuatro parámetros: en primera posición se encuentra “Ch”, que identifica a los niños, y “Ad”, que identifica a los adultos. En segunda posición, “M” indica masculino y “F” indica femenino. Los siguientes valores son la edad del participante, y el último número diferencia a dos casos cuando repiten las descripciones anteriores.

Tabla 1. Género, edad y zona de procedencia de los niños participantes

No	Género y edad	Notación	Zona de procedencia
1	Masculino, 6 años	Ch.M.06.1	San José
2	Masculino, 11 años	Ch.M.11.1	San José
3	Femenino, 8 años	Ch.F.08.1	San José
4	Femenino, 8 años	Ch.F.08.2	Alajuela
5	Femenino, 12 años	Ch.F.12.1	Alajuela

Esta sesión fue realizada en el estudio de grabación de la escuela de Ciencias de la Comunicación Colectiva y se contó con el consentimiento informado de los padres de los niños y niñas participantes, documento que fue elaborado teniendo como base el formato que dicta la Universidad de Costa Rica para este efecto.

De la misma forma se realizaron dos sesiones con jóvenes adultos estudiantes del nivel de bachillerato de las carreras de educación especial e ingeniería eléctrica. Esta grabación fue realizada durante el primer semestre del año 2019.

Tabla 2. Género, edad y zona de procedencia de los adultos participantes

No	Género y edad	Notación	Zona de procedencia
1	Masculino, 17 años	Ad.M.17.1	Heredia
2	Masculino, 18 años	Ad.M.18.1	Heredia
3	Masculino, 20 años	Ad.M.20.1	Osa
4	Femenino, 20 años	Ad.F.20.1	San José
5	Femenino, 21 años	Ad.F.21.1	Heredia
6	Femenino, 21 años	Ad.F.21.1	San José
7	Femenino, 21 años	Ad.F.21.3	Limón
8	Femenino, 23 años	Ad.F.23.1	Guanacaste
9	Femenino, 26 años	Ad.F.26.1	San José

Es necesario en este punto enfatizar que el carácter del presente estudio es exploratorio, ya que la cantidad de participantes y la muestra inicial de datos con los que se trabajó es la primera para un análisis como el presente. La limitante de trabajar con muestras grandes de participantes, en especial niños, es particularmente relevante para la generalización de la información. Sin embargo, como se ha mostrado para otros idiomas, los estudios de carácter exploratorio dan indicios valiosos para la toma de decisiones y los ajustes necesarios en el desarrollo tecnológico, al hacer notar condiciones particulares que pueden estar siendo desatendidas por los sistemas actuales.

3.4 Metodología de grabación

Para el diseño de la base de datos se desarrolló una estrategia de interacción con los niños y niñas participantes en la que se utilizaron instrumentos formales y no formales propios de la valoración del lenguaje oral, con el objetivo principal de recolectar las muestras de voz en lugar de valorar la articulación de cada fonema o la construcción de oraciones.

Inicialmente se utilizó el test de articulación, mismo que contempla la emisión de los sonidos iniciales, medios y finales de los fonemas del alfabeto en español, al mostrar diversas imágenes que sugieren a la persona participante el uso del habla espontánea para indicar de qué se trata.

Así, por ejemplo, se muestra un dibujo con la imagen de una media y se le pregunta al niño o niña “¿qué es esto?”, con el propósito de valorar en la hoja de respuesta si se dio alguna omisión, sustitución o distorsión del sonido “M” en el caso específico de la palabra media. No obstante, en el caso de este estudio, como se indicó con anterioridad, el test en cuestión se seleccionó por el hecho de que valora los diversos sonidos de los fonemas en español y no para evaluar la calidad de la articulación de los mismos.

Se realizó también la grabación de palabras por grupos semánticos con el propósito de contar con un banco de información por categorías de alto uso en el lenguaje infantil. Dichos grupos semánticos fueron definidos tomando como base tanto actividades de la vida diaria como actividades que forman parte del currículo escolar, y que son propias de la primera infancia (colores, animales, alimentos) (Morales, 2010).

Se grabaron también palabras que contemplaran las categorías semánticas antes señaladas, tanto en singular como en plural y finalmente la construcción de oraciones

según temas de interés de las personas menores participantes, con el fin de generar bancos relacionados a temas de alto agrado que, por ende, cuentan con impacto emocional y permiten registrar las inflexiones en la voz. Algunas de estas palabras se recolectaron utilizando temas libres donde se registró información sobre: mitología, en especial de hadas y ninfas; Pokémon; animales, particularmente perros; viajes o paseos.

Al hablar de las estrategias usadas para la toma de las muestras de voz en la interacción con menores de edad es importante considerar que se requiere de conocimiento en el área de desarrollo infantil, no solo por la forma en que es propicio relacionarse con los participantes, sino también porque se debe tener conocimiento de los aspectos de percepción y atención esperados según la edad de cada niño o niña, para poder aprovechar al máximo los insumos que se necesiten recopilar.

Otro aspecto fundamental en el trabajo realizado con los y las menores se centró en propiciar un ambiente agradable y poco intimidante para procurar que se sintieran cómodos, para ello se tomó el tiempo de explicar qué era y cómo funcionaba un estudio de grabación, así como permitir a las y los participantes estar en contacto con los instrumentos de grabación de manera exploratoria previo a la sesión formal de grabación.

Dentro de las estrategias usadas directamente en la interacción con los niños y niñas participantes, por una parte, se usó material con buen contraste visual, lo cual es necesario para garantizar una mejor percepción visual y también apoyar la atención en el material. Por otra parte, se utilizaron juegos de competencias para la pronunciación de palabras en donde tanto la mediadora del proceso como los participantes pronunciaban las diferentes palabras.

En la dinámica de trabajo durante la grabación, fue importante alternar la recolección de datos con espacios de interacción no formal, en los cuales se aprovechó para registrar información sobre temas libres de interés, tomar descansos y usar el reforzamiento positivo de carácter verbal.

La cantidad de vocales para cada uno de los hablantes se muestra en la Tabla 3. En esta tabla se reporta la cantidad mínima de vocales por cada uno de los participantes. Si bien en la mayoría de los casos se cuenta con un número mayor de emisiones de los sonidos, no fue posible analizar algunos de estos porque no se detectaron con la herramienta *EasyAlign*, o del todo la palabra se registró ininteligible.

Tabla 3. Cantidad mínima de vocales grabadas por cada uno de los participantes, tanto niños como adultos

Vocal	Cantidad
a	135
e	56
i	29
o	68
u	21

3.5 Extracción de información acústica

Cada frase o palabra de la base de datos pronunciada por los participantes fue segmentada en los fonemas correspondientes. Adicionalmente, la duración, los dos primeros

formantes y el tono de cada una de las cinco vocales fue extraído usando el método de autocorrelación, con el programa *Praat* (Goldman y Shwab, 2014), como se ilustra en la Figura 1.

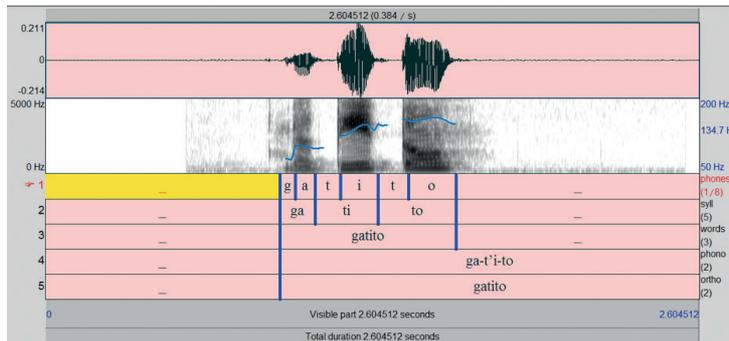


Figura 1. Interfaz de *Praat* y segmentación realizada con *EasyAlign*

Para lograr la gran cantidad de análisis realizados, se desarrollaron rutinas de programación en este sistema, con las cuales se automatizó el proceso de extracción de información. Con estos programas, se puede tener como entrada un archivo de audio y su correspondiente transcripción en texto y, como salida, un conjunto de archivos generados automáticamente por *Praat*, los cuales se detallan a continuación:

- Duraciones (cinco archivos, uno por cada vocal): en estos archivos, gracias a la segmentación automática generada por el programa, se escribe en cada línea la vocal correspondiente, la marca de tiempo de su inicio y la marca de tiempo de su finalización. Con la resta de ambos de ambas marcas de tiempo se calcula la duración de cada una.
- Formantes (cinco archivos, uno por cada vocal): en estos archivos, generados a partir de lo programación realizada, se escribe en cada línea la vocal correspondiente y los valores de los primeros cuatro formantes.
- Tono (cinco archivos, uno por cada vocal): en estos archivos se obtiene una línea por cada una de las vocales analizadas, en los cuales se registra el tono medio de cada una de las vocales.

De esta manera, aunque el proceso logra automatizar la segmentación de cada uno de los sonidos y la obtención de las características de los sonidos de interés (las vocales), el proceso requiere de un tiempo considerable de cómputo (del orden de horas), para extraer toda la información. Posteriormente, con los archivos de datos generados se obtuvieron las estadísticas o gráficos correspondientes reportados en la sección de resultados, de acuerdo con los grupos de edad, género, parámetro y vocal en particular.

Como ilustración, en la Figura 2 se muestra una gráfica generada con ayuda de una herramienta de hoja de cálculo, con la dispersión de los dos primeros formantes de la vocal /i/, que compara los que produjo una niña con los de una mujer adulta. En uno de los ejes se muestra el valor en Hertz de la ubicación del primer formante para cada una de las vocales grabadas de los participantes, mientras que en el otro eje se muestra el valor en Hertz del segundo formante. Sobre esto, se ha estudiado ampliamente el hecho de que los dos primeros formantes pueden caracterizar las vocales del español (Martínez Celdrán, 1995).

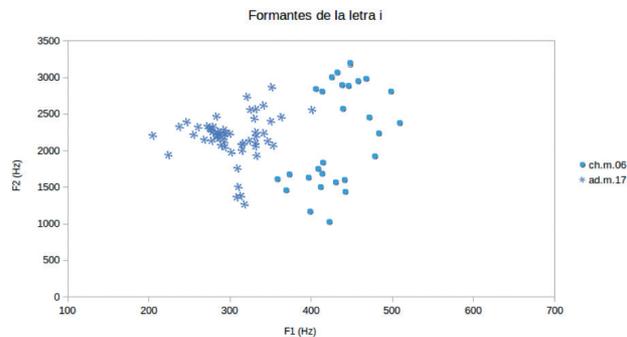


Figura 2. Ubicación de los dos primeros formantes de la letra /i/ producida por una niña y una mujer adulta

La duración reportada en el presente trabajo es calculada a partir de los límites establecidos en los fonemas por la segmentación realizada con el programa *EasyAlign* de *Praat*.

4. Resultados

En esta sección se presentan los resultados de los análisis acústicos realizados sobre las vocales del habla de las personas menores participantes. Estos son similares a los presentados en otros idiomas, como en el análisis realizado por Gerosa et al. (2007) para los idiomas inglés e italiano. Estos análisis se llevaron a cabo para lograr una mejor comprensión de los cambios que ocurren en el habla producidos por niños y niñas de diversas edades, y para establecer una comparación con el habla de los adultos. Los parámetros considerados en el análisis son la duración de las cinco vocales, sus formantes y el tono. Se debe hacer énfasis en el hecho de que el interés principal es el análisis de la emisión de las vocales en los menores, y que los resultados de los adultos se presentan solamente con fines comparativos.

La información se presenta principalmente en diagramas de caja, los cuales permiten la visualización de la distribución de todos los valores en los parámetros.

4.1 Duración de las vocales

Las Figuras 3 a 7 muestran los resultados de las duraciones de las vocales de cada uno de los participantes, en forma de diagramas de caja. Por ejemplo, en la Figura 3, se presenta la duración de la vocal /a/ para todos los niños y adultos registrados. El rango más amplio de duración provino del niño varón de 6 años (Ch.M.06.1), con valores entre 0.02 y 0.68 segundos. Las tres niñas presentan una distribución de duración similar en esta vocal y, a excepción de dos mujeres adultas, todos los adultos presentan un rango de duración similar, con una distribución que muestra valores más cortos en la pronunciación en todos los casos, hasta incluso llegar a valores entre 0.02 y 0.19 segundos en uno de los casos, muy por debajo del rango de los niños.

La mediana de los valores es claramente mayor en el caso de los niños, la cual se ubica sobre 0.15 segundos, con respecto a los adultos masculinos y femeninos, cuya mediana está bajo 0.1 segundos. Se encuentran notables diferencias entre los niños varones y los varones adultos, al compararlas con las diferencias existentes entre las niñas y las mujeres adultas.

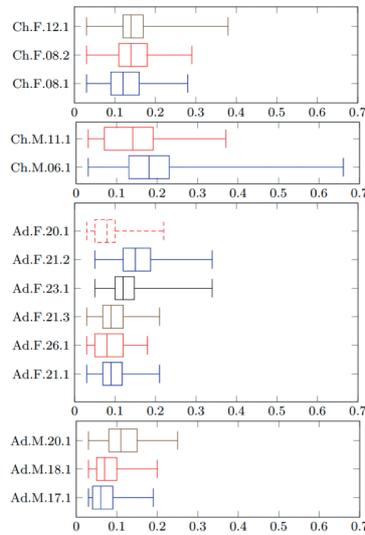


Figura 3. Diagramas de la duración de la vocal /a/ (en segundos) para los niños comparados con adultos y separados por género

Estos diagramas de caja permiten observar no solamente el rango de las duraciones (del valor mínimo al máximo), sino el rango principal donde están concentrados la mayoría de los valores (los cuartiles centrales).

Los resultados de una media mayor de duración se pueden observar en la vocal /e/ (Figura 4), y vocal /o/ (Figura 6). En ambos casos, los rangos menores corresponden a los varones adultos, concentrados entre 0.02 y 0.18 segundos, mientras que los mayores a los niños varones, con rangos que van desde los 0.02 segundos y superan los 0.4 segundos.

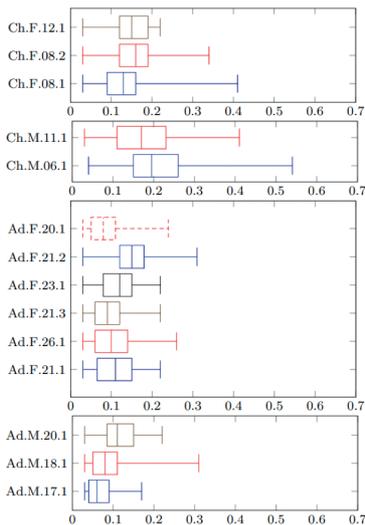


Figura 4. Diagramas de la duración de la vocal /e/ (en segundos) para los niños comparados con adultos y separados por género

También es importante destacar que todos los diagramas mostrados en estas figuras tienen la misma escala en el eje de valores, de manera que es posible una comparación visual. Por ejemplo, las niñas Ch.F tienen valores más homogéneos que los niños varones Ch.M en las primeras dos vocales: <a> y <e>.

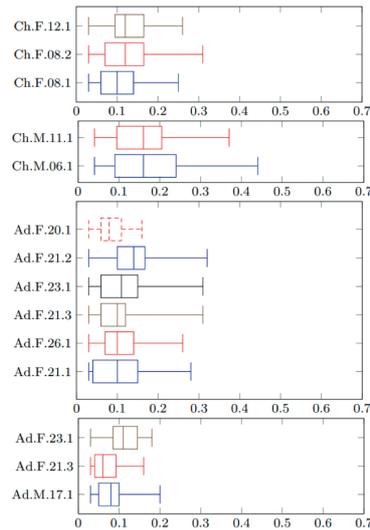


Figura 5. Diagramas de la duración de la vocal /i/ (en segundos) para los niños comparados con adultos y separados por género

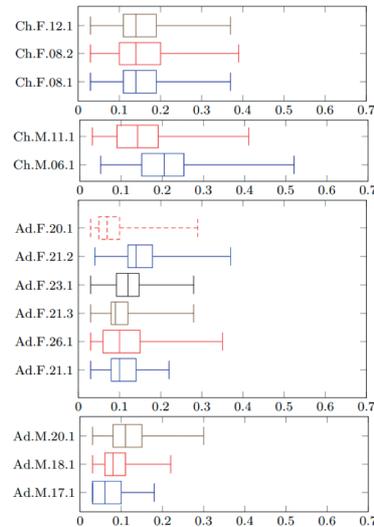


Figura 6. Diagramas de la duración de la vocal /o/ (en segundos) para los niños comparados con adultos y separados por género

Las Figura 5 y la Figura 6 permiten observar que las vocales <i> y <o> son las que tienen la duración más homogénea con respecto a las demás, tanto en los niños como en los adultos. En el caso de la <i>, se presenta el rango de duraciones menor para Ch.M.06.1, con valores que no superan los 0.45 segundos, a diferencia de los valores de la <a> que se acercan a los 0.7 segundos.

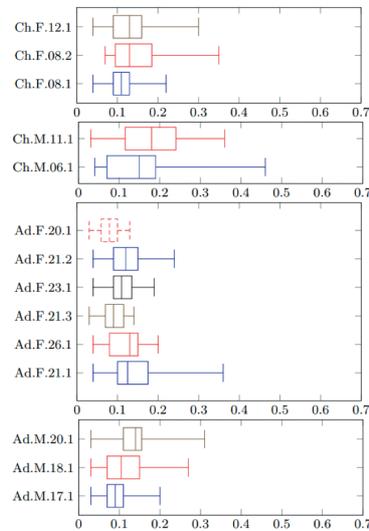


Figura 7. Diagramas de la duración de la vocal /u/ (en segundos) para los niños comparados con adultos y separados por género

La Figura 7 muestra que la duración de la vocal <u> es la que más diferencias presenta entre los hablantes, con un rango particularmente estrecho de valores para Ad.F.20.1 (entre 0.015 y 0.12 segundos) y muy amplio para Ch.M.06.1 (entre 0.03 y 0.47 segundos). Esto puede señalar la duración de esta vocal como uno de los mayores diferenciadores entre la pronunciación de vocales de niños y de adultos.

En todos los casos, los rangos de duración de vocales en niños son mayores que en adultos, con las diferencias más marcadas en los niños varones. Para el caso de la vocal /u/, el rango y medio de duración de los varones adultos se asemeja al de las niñas. Sin embargo, debe tomarse en consideración, de acuerdo con lo presentado en la Tabla 2, que se tiene una cantidad mucho menor de esta vocal en comparación con las demás.

Otra información relevante que se desprende de las figuras anteriores es que no se encuentra homogeneidad en cuanto a las duraciones a lo largo de las vocales y de los participantes. Es decir, si bien Ad.M.20.1 tiende a tener mayor duración en la /a/, /o/, /u/, no es así con la /e/ y con la /i/, como se observa en las Figuras 4 y 5.

Como una observación general, los rangos y medianas tienden a ser mayores entre los niños varones que entre las niñas. En algunos casos, como en la vocal /i/ y en la /u/, los valores de duración de las niñas se asemejan más a las mujeres adultas que a los niños varones.

4.2 Análisis de formantes

Los formantes son picos de intensidad que se presentan en el espectro de una señal. En la literatura se reconoce la relación que existe entre los cambios en los formantes de las vocales con las características y la forma del tracto vocal. Para realizar análisis comparativos, usualmente se estudia el rango y la variación de la posición de los dos primeros formantes, como se ilustra en la Figura 8 para las cinco vocales de uno de los participantes.

Por su parte, la Figura 9 muestra la carta de formantes, calculada como la posición media de los dos primeros formantes para dos niñas y dos mujeres adultas del conjunto de datos, utilizando la misma escala en ambos ejes. Esta representación ha sido mostrada, por ejemplo, por Martínez Celdrán (1995) para caracterizar el habla de diferentes grupos.

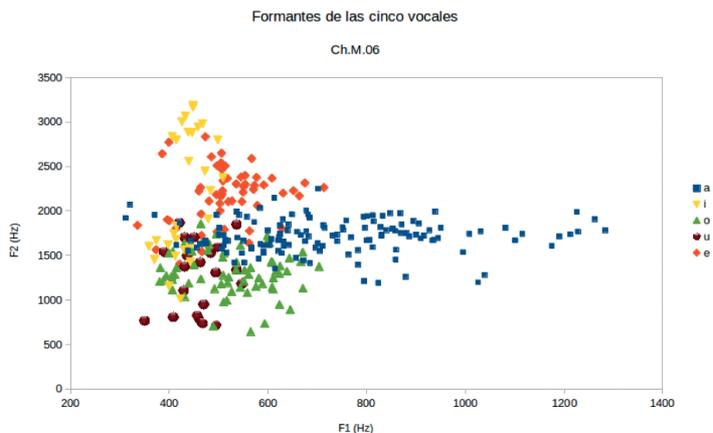


Figura 8. Campo de dispersión de los formantes de las vocales producidas por Ch.M.01

Las niñas tienen patrones variables en el polígono de los formantes. Por su parte, las adultas muestran un rango más amplio con respecto a los formantes y preservan una forma más homogénea en este polígono.

En términos del mismo espacio de formantes, se puede hacer una observación similar en la Figura 10, para el caso de los hablantes masculinos. Los polígonos para los dos niños varones tienen diferencias significativas entre ellos y difieren de los dos adultos. En este caso, también los polígonos de los adultos tienen una forma mucho más homogénea. Es importante notar que el niño de 11 años tiene un espacio de formantes en vocales más semejante al de los adultos que el niño de 6 años.

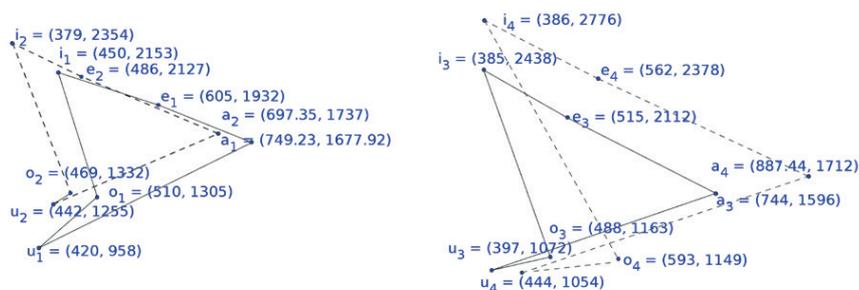


Figura 9. Carta de las vocales en el espacio de los dos primeros formantes, para el caso de hablantes femeninas. Izquierda: Niñas. Derecha: Adultas

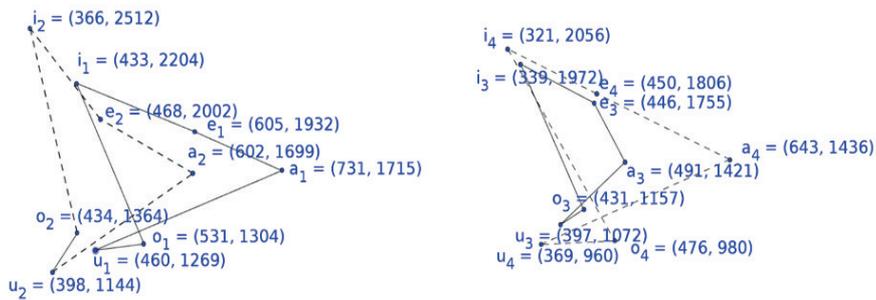


Figura 10. Carta de las vocales en el espacio de los dos primeros formantes, para el caso de hablantes masculinos. Izquierda: Niños. Derecha: Adultos

4.3 Tono de las vocales

Para el análisis del tono de la vocal, presentamos los diagramas de caja para los datos correspondientes a dos tipos de vocales emitidas para los niños y los adultos, las cuales son las más representativas del estudio. El primero es la vocal abierta /a/, en la Figura 11. Con la excepción de una mujer adulta, el rango de tono para esta vocal, cercano al intervalo 100 Hz a 275 Hz, es evidentemente más amplio en todos los niños que en los adultos, independientemente de su género. Las diferencias con el rango y el amplio rango de valores para hombres adultos son notables, pues en estos últimos los valores tienen un rango entre 100 y 175 Hz aproximadamente, y se concentran en su mayoría en valores cercanos a los 100 Hz.

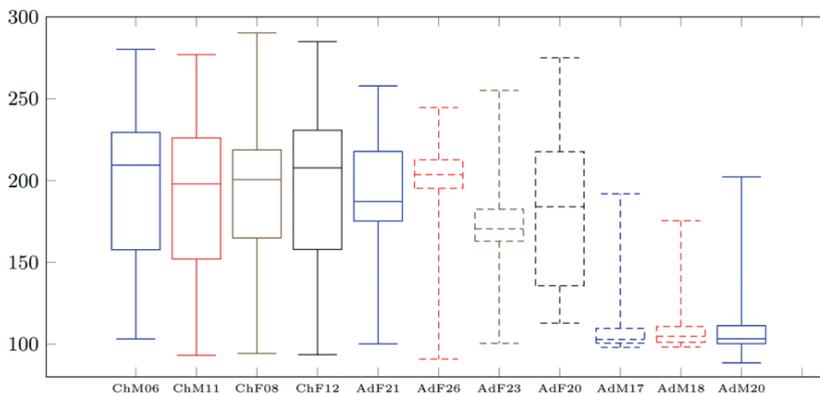


Figura 11. Diagramas de caja del tono de la vocal /a/ por cada hablante

Las notables diferencias de tono y rango se observan en las vocales /e/ (Figura 12), /o/ (Figura 14), /u/ (Figura 15), las cuales distinguen el tono de los hombres adultos del de los niños y mujeres de todas las edades analizadas. En todos los casos, el rango de tono de las vocales de los niños es el rasgo distintivo más relevante de estos datos, con una mayoría de valores centrados entre los 175 y 225 Hz, cercano al de las mujeres adultas, aunque en estas está más concentrado entre 210 y 225 Hz. En el caso de la vocal /i/ (Figura 13), en dos de los adultos (uno hombre y la otra mujer) presentaron rangos que son equiparables a los de los niños, aunque el tono medio, sobre todo en el caso del hombre, sigue siendo menor y muy concentrado cerca de los 100 Hz.

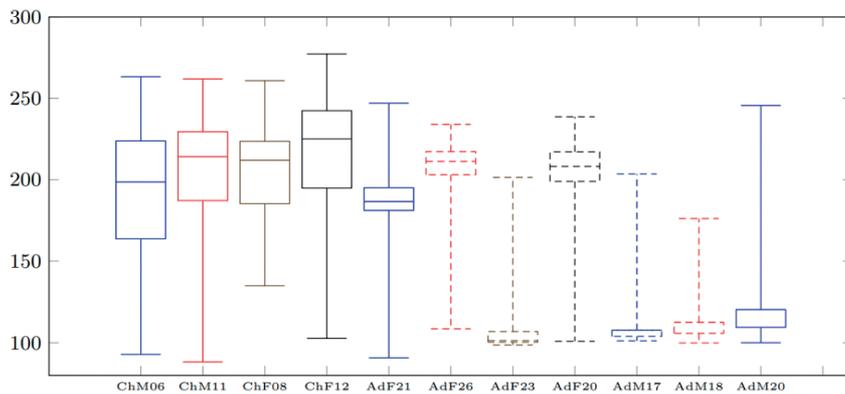


Figura 12. Diagramas de caja del tono de la vocal /e/ por cada hablante

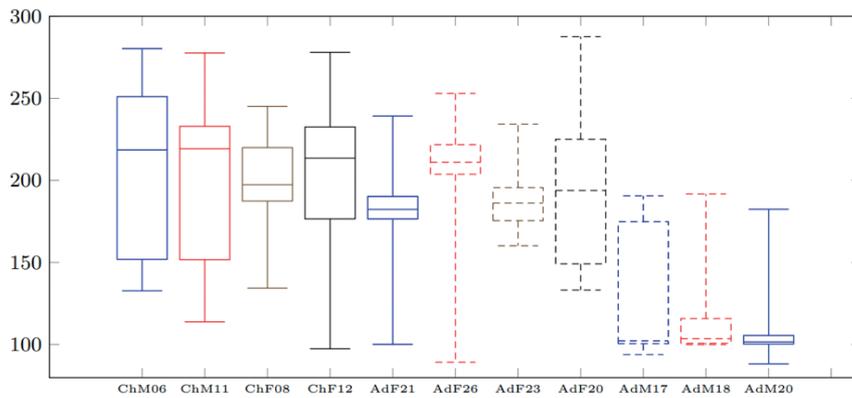


Figura 13. Diagramas de caja del tono de la vocal /i/ por cada hablante

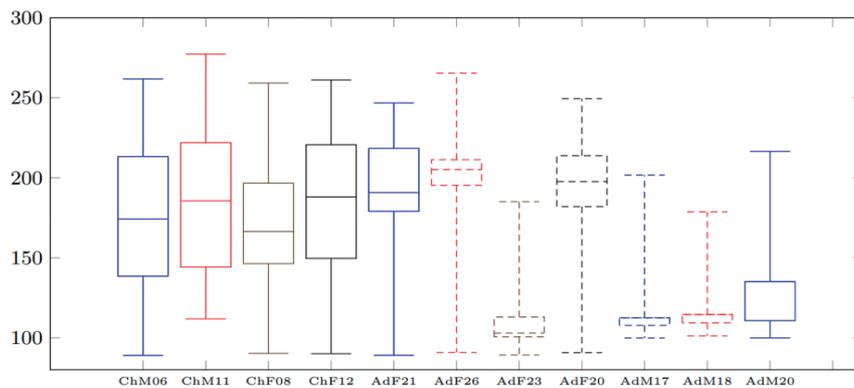


Figura 14. Diagramas de caja del tono de la vocal /o/ por cada hablante

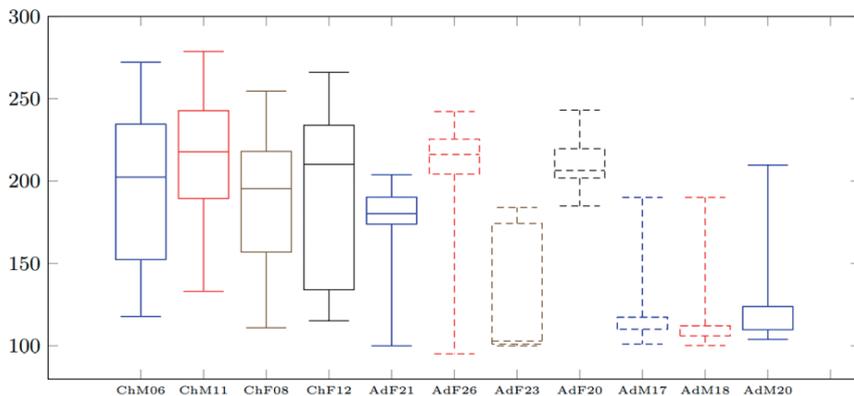


Figura 15. Diagramas de caja del tono de la vocal /u/ por cada hablante

5. Conclusiones

En este artículo, se presentaron los resultados de un análisis acústico realizado en las vocales del habla en niños costarricenses, a partir de una base de datos generada para tal fin. El análisis se centró en la duración de las vocales, el tono y los formantes. Los participantes fueron niños con edades entre los 6 y los 12 años, quienes pronunciaron un conjunto de palabras especialmente seleccionadas de acuerdo con su edad y su desarrollo.

El análisis muestra que los niños tienden a pronunciar las cinco vocales del español con mayor duración, con mayores diferencias entre los adultos varones y los niños varones. En los adultos, independientemente de su género, hay una duración más homogénea en todos los casos, en comparación con los niños.

Los polígonos de formantes muestran mayores variaciones entre los niños, con patrones muy diferentes en hombres y mujeres. En comparación con los adultos, parece haber una relación con la edad, donde los niños mayores presentan los polígonos que son más similares a los de los adultos.

En cuanto a la relación niños y adultos con respecto al tono de las vocales, las diferencias en la mayoría de ellas son notables en cuanto al tono medio y al rango: tanto los niños como las niñas tienden a pronunciar las cinco vocales analizadas en un rango de tonos mayor que el de los adultos y con una media más elevada que la de varones adultos. Por su parte, los rangos de tono son más parecidos entre el grupo de menores y las mujeres adultas.

Todos estos resultados son de valor para el conocimiento de las características del habla en esta población, y permiten futuras investigaciones en áreas como el reconocimiento de voz dependiente del género y la edad, así como la implementación de algoritmos para mejorar el rendimiento de los reconocedores automáticos del habla para esta variante particular del español.

Bibliografía

Calvo Shadid, A. (2008). Las semiconsonantes y semivocales en los diptongos del español: propuesta de análisis fonológico. *Revista de Filología y Lingüística de la Universidad de Costa Rica*, 34(2), 107-142.

- Coto-Jiménez, M., Goddard-Close, J. y Martínez-Liconá, F. M. (2014a). Quality Assessment of HMM-Based Speech Synthesis Using Acoustical Vowel Analysis. En A. Ronzhin, R. Potapova y V. Delic (Eds), *Speech and Computer* (pp. 368-375). Suiza: Springer.
- Coto-Jiménez, M., Martínez-Liconá, F. M. y Goddard-Close, J. (2014b). Acoustic Vowel Analysis in a Mexican Spanish HMM-based Speech Synthesis. *Research in Computing Science*, 86, 53-62.
- Delgado, C. (2014). *Viajando a Ítaca por los mares cuantitativos. Manual de ruta para investigar en grado y postgrado*. Salamanca, España: Ediciones Amarú.
- Díaz, S., Cisternas, P. y López, I. (2015). Características acústicas de las vocales del español de Chile producidas por sujetos residentes en la ciudad de Santiago. *Revista Chilena de Fonoaudiología*, 14, 92-102.
- Gerosa, M., Giuliani, D. y Brugnara, F. (2007). Acoustic variability and automatic recognition of children's speech. *Speech Communication*, 49(10-11), 847-860.
- Goldman, J. P. y Schwab, S. (2014). Easyalign Spanish: An (Semi-)Automatic Segmentation Tool Under Praat. En Y. Martín Congosto, M. L. Montero Curiel y A. Salvador Plans (Eds.), *Fonética Experimental, Educación Superior e Investigación* (5 vol.) (629-640). Madrid: Arco Libros. doi: <https://doi.org/10.5167/uzh-107124>.
- Hagen, A., Pellom, B. y Cole, R. (2003). Children's speech recognition with application to interactive books and tutors. *2003 IEEE Workshop on Automatic Speech Recognition and Understanding* (pp. 186-191). doi: 10.1109/ASRU.2003.1318426.
- Hernández Sampieri, R., Fernández Collado, C. y Baptista Lucio, M. P. (2010). *Metodología de la Investigación*. (5 ed.) México D.F.: McGraw-Hill.
- Leinonen, T. N. (2010). *An acoustic analysis of vowel pronunciation in Swedish dialects*. (Tesis de doctorado). Universidad de Groningen.
- Li, Q. y Russell, M. J. (setiembre, 2002). An analysis of the causes of increased error rates in children's speech recognition. Conferencia llevada a cabo en la *7th International Conference on Spoken Language Processing*. Interspeech: Denver, EE.UU.
- Marín Gálvez, R. (1995). La duración vocálica en español. *ELUA: Estudios de Lingüística Universidad de Alicante*, (10), 213-226.
- Martínez Celdrán, M. (1995). En torno a las vocales del español: análisis y reconocimiento. *Estudios de fonética experimental*, 7, 195-218.
- Martínez-Liconá, F. M., Goddard-Close, J. y Coto-Jiménez, M. (2013). Acoustic Analysis of Spanish Vowels in Emotional Speech. En C. Manfredi (Ed), *Models and Analysis of Vocal Emissions for Biomedical Applications. 8th International workshop* (pp. 235-238). Firenze: Firenze University Press.
- Morales, M. (2010). *De los sonidos a las palabras: métodos y técnicas para la estimulación y corrección del lenguaje oral I*. San José, Costa Rica: EUNED.
- Petrović-Lazić, M., Babac, S., Vuković, M., Kosanović, R. y Ivanković, Z. (2011). Acoustic voice analysis of patients with vocal fold polyp. *Journal of Voice*, 25(1), 94-97.

- Potamianos, A. y Narayanan, S. (2003). Robust recognition of children's speech. *IEEE Transactions on speech and audio processing*, 11(6), 603-616.
- Safavi, S., Najafian, M., Hanani, A., Russell, M. J., Jancovic, P. y Carey, M. J. (setiembre, 2012). Speaker recognition for children's speech. Conferencia llevada a cabo en la *13th Annual Conference International Speech Communication Association*. Interspeech: Portland, EE. UU.
- Salamanca Gutiérrez, G. F. y Valverde San Martín, A. L. (2009). Prestigio y estigmatización en variantes anteriorizadas y posteriorizadas de las vocales del español de Chile. *Literatura y lingüística*, 20, 125-140.
- Schwab, S. (2015). Las variables temporales en el español de Costa Rica y de España: un estudio comparativo. *Revista de Filología y Lingüística de la Universidad de Costa Rica*, 41(1), 127-139.
- Serrano, J. (2006). En torno a las vocales caedizas del español mexicano: una aproximación sociolingüística. En P. M. Butragño (Ed.), *Líderes lingüísticos: Estudios de variación lingüística* (pp. 37-59). Ciudad de México: El colegio de México.
- Skodda, S., Visser, W. y Schlegel, U. (2011). Vowel articulation in Parkinson's disease. *Journal of voice*, 25(4), 467-472.
- Soto Barba, J. (2007). Variación del F1 y del F2 en las vocales del español urbano y rural de la provincia de Ñuble. *RLA. Revista de Lingüística Teórica y Aplicada*, 45(2), 143-165.
- You, H., Alwan, A., Kazemzadeh, A. y Narayanan, S. (setiembre, 2005). Pronunciation variations of Spanish-accented English spoken by young children. Conferencia llevada a cabo en la *9th European Conference on Speech Communication and Technology*. Interspeech: Lisboa, Portugal.