



**UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE GUAYAQUIL**

**TRABAJO DE GRADO PREVIO A LA OBTENCIÓN DEL
TÍTULO DE:
INGENIERO DE SISTEMAS**

**CARRERA:
INGENIERÍA DE SISTEMAS**

**TEMA:
"RECONOCIMIENTO AUTOMÁTICO DE VOZ APLICADO A LA
MEJORA EN EL PROCESO DE APRENDIZAJE DE LECTURA EN
NIVEL ESCOLAR"**

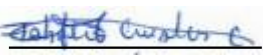

**AUTOR:
FABIOLA LISSETT CEVALLOS CORREA.**

**TUTOR:
MONICA DANIELA GOMEZ RIOS Msg.**

**Abril 2021
GUAYAQUIL-ECUADOR**

DECLARATORIA DE RESPONSABILIDAD

Yo, Fabiola Cevallos Correa , declaro que los conceptos y análisis desarrollados y las conclusiones del presente trabajo son de exclusiva responsabilidad del autor.

 Firma del autor Nombre: Fabiola Cevallos Correa CI.0930870779	 Firma del tutor Nombre: Mónica Gómez Ríos- CI.0104606777
---	--

“Reconocimiento automático de voz aplicado a la mejora en el proceso de aprendizaje de lectura en el nivel escolar”

Mónica Gómez R.¹[0000-1111-2222-3333] y Fabiola Cevallos²[1111-2222-3333-4444]

¹Universidad Politécnica Salesiana, Guayaquil, Ecuador
mgomezr@ups.edu.ec, y fcevallosc@est.ups.edu.ec

Resumen. El presente artículo realiza el estudio de los algoritmos y los métodos que se utilizan para el reconocimiento automático de la voz. La importancia de su aplicación reside en la mejora del proceso de aprendizaje tomando en cuenta características como la pronunciación, comunicación y habilidades lingüísticas en mejora de la precisión y buen uso del habla. Se realizó una búsqueda de los diferentes estudios relacionados con el objetivo de analizar el algoritmo más utilizado y el método que lo complementa para así poder encontrar las características más relevantes e inconvenientes en su utilización. Siendo así, se encontró que uno de los algoritmos más utilizados es DNN (redes neuronales profundas) y HMM (modelo oculto de Markov). De esta revisión pudimos observar que el uso de reconocimiento de la voz para el aprendizaje ha sido experimentado desde edades tempranas en niños, adolescentes y adultos con resultados relevantes, siendo uno de los inconvenientes principales para su funcionamiento, el ruido.

Palabras clave: Reconocimiento de voz, Algoritmos, DNN (Redes Neuronales Profundas), HMM (Modelo Oculto de Markov), algoritmo de Viterbi, algoritmo de Baum-Welch, aprendizaje.

1. Introducción

La dislalia infantil es la mala pronunciación y la falta de articulación de los fonemas este es un trastorno muy frecuente en los niños, una de las técnicas o metodologías para mejorar o identificar este problema es el reconocimiento automático de la voz, el cual está siendo utilizado como una técnica de apoyo a la mejora del aprendizaje de la lectura en nivel escolar. Nuestro interés se enfoca en analizar los distintos algoritmos y métodos que predominan para su funcionamiento y conocer cómo esta tecnología puede ser utilizada en el aprendizaje de niños con problemas de imperatividad o simplemente niños con problema de lenguaje. La Inteligencia Artificial es una ciencia computacional que crea una máquina para representar las características y capacidades similares al del ser humano [16], una de estas capacidades es el reconocimiento automático de la voz RAH, el cual que es uno de los tipos de inteligencia artificial que intenta establecer la comunicación entre el computador y el hombre al identificar y procesar su voz. La identificación de la voz no es una tarea fácil, ya que depende de diferentes aspectos de cada persona como: velocidad al hablar, esfuerzo vocal, idiomas o factores ambientales. Posee varias utilidades, como medio de seguridad o esta-

blecimiento de comandos al autenticar la identidad de la persona que habla en un sistema [12], también sirve como herramienta de aprendizaje.

La metodología para este análisis descriptivo fue mediante una revisión bibliográfica en los buscadores más destacados en el área tomando en cuenta diferentes criterios de inclusión y exclusión. Las aplicaciones desarrolladas con este fin utilizan diferentes tipos de dispositivos, por lo general con un robot. El funcionamiento de los algoritmos se da mediante un conjunto ordenadores que realizan operaciones sistemáticas y permiten hacer un cálculo y hallar la solución de un tipo de problema. Entre los algoritmos más utilizados del reconocimiento de la voz de acuerdo a esta revisión están: a) RNP (Redes neuronales profundas) que es un tipo de red artificial basada en un conjunto de unidades neuronales, que tiene una forma similar a las neuronas del cerebro biológico [1]. b) el algoritmo LSTM-RNN, que realiza un aprendizaje más profundo, no solo porque puede procesar puntos de datos individuales, sino también secuencias completas de datos permitiendo hallar la secuencia más probable de estados ocultos, c) Baum-Welch busca la similitud de las características de la voz encontrando los parámetros desconocidos en una base de datos, d) [18].Viterbi, realiza el aprendizaje profundo que admiten transformaciones no lineales múltiples e iterativas de datos [19], y e) el algoritmo RTISI-LA (Inversión de espectro iterativo en tiempo real con mira hacia el futuro) que calcula el espectro de la señal a corto plazo obtenida por superposición. En cuanto a los métodos con los cuales funciona cada algoritmo tenemos: a) HMM, es un modelado estadístico más utilizado para series de tiempo y secuencias, b) VTLN calcula los factores de distorsión de frecuencia, la cual es una técnica que elimina frecuencias de distorsión mejorando el reconocimiento de la voz. Este estudio se divide en las siguientes secciones: a) descripción metodológica, b) descripción de los algoritmos y métodos utilizados en el reconocimiento de la voz, c) ventajas y desventajas del uso del reconocimiento de la voz en el aprendizaje, d) usuarios potenciales, e) resultados y f) conclusiones.

2. Metodología

El presente estudio realiza una revisión bibliográfica descriptiva acerca de las principales características en las que ha influenciado el reconocimiento de voz para el aprendizaje. Para ello nos hemos respondido a las siguientes preguntas de investigación que se pueden observar en la tabla 1.

Pregunta 1	¿Cuáles son los algoritmos y métodos utilizados en el reconocimiento de la voz aplicado al aprendizaje?
Pregunta 2	¿Cuáles son las Ventajas e inconvenientes de los algoritmos de reconocimiento de la voz en su aplicación?
Pregunta 3	¿Quiénes son los usuarios potenciales con los que se ha trabajado en el reconocimiento de la voz para el aprendizaje?

Tabla 1. Resultado de la búsqueda

La búsqueda se realizó en las bibliotecas más comúnmente utilizadas, tomando en cuenta los trabajos de los últimos 5 años y relacionados con el aprendizaje mediante la siguiente consulta que se realizó en Exel de la matriz en la que consiste en ingresar palabras claves, se obtuvo como resultado una ecuación en la que podemos buscar y realizarlas distintas consultas se realizó, en distintas bibliotecas. Siendo así, los resultados obtenidos se detallan en la Fig 1.

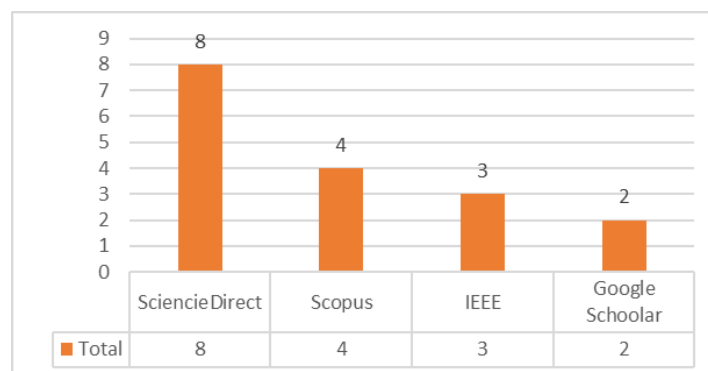


Fig 1. Resultado de la búsqueda

Como se puede observar en la Fig 1, la biblioteca donde más trabajos se encontraron fue Science Direct, seguida de Scopus, IEEE y Google Scholar, con un total de 17 artículos relacionados.

3. Algoritmos y métodos utilizados en Speed Recognitio

El reconocimiento de la voz lo obtenemos mediante la aplicación de distintos algoritmos, a continuación, en la Tabla 2, se detallan los más utilizados de acuerdo con los trabajos revisados [1][18] [8],

Tabla 2: Algoritmos de reconocimiento de voz.

No	Baum-Welch s	Viterbi	DNN	LSTM-RNN	Algoritmos de aprendizaje profundo
1	-	-	-		
2			-		
3			-		
4		-			
5			-		
6		-	-	-	
7		-	-		
8					-
9			-		
10					-
11			-	-	
12					
13			-	-	
14				-	
15	-		-		
16			-		
17		-	-		

Al relacionar los diferentes algoritmos, el algoritmo de entrenamiento de Baum-Welchs busca la similitud de la característica de la voz en cuanto a la intensidad y timbre del habla. A diferencia del algoritmo Viterbi que busca la ruta óptima del reconocimiento de una secuencia de voz en particular. Por otra parte, el algoritmo DNN (Redes neuronales profundas) realiza cálculos progresivamente complejos sobre las señales de voz y las clasifica; y el algoritmo LSTM- RNN reconoce una señal bidireccional que captura información desde cualquier lugar de la secuencia de la voz [11]. Todos estos algoritmos aportan para el desarrollo de aplicaciones que realicen reconocimiento de la voz, siendo el más utilizado, el algoritmo de DNN [8].

4. Métodos que utiliza el reconocimiento de la voz

Un algoritmo trabaja junto con un método, el cual lo complementa y consiste generalmente en una serie de sentencias para llevar a cabo una acción [14]. Entre los más comúnmente utilizados de acuerdo con esta revisión podemos observarlos en la Tabla 2.

Por otro lado, al relacionar los distintos métodos, el MFCC busca la similitud de la característica de la voz, separa la señal de voz y la de ruido, a diferencia HMM modelo oculto de Markov que extrae los parámetros desconocidos del reconocimiento de la voz. Por otra parte, RTISI-LA, construye la señal reproducida en el dominio del tiempo a diferencia del VTLN trasforma la voz y elimina los factores de distorcion. Entre

los métodos más utilizado está el HMM, el cual realiza un reconocimiento de voz en secuencia extrayendo los parámetros de la voz.

Tabla 3. Métodos de reconocimiento de la voz.

No	(MFCC)	HMM	RTISI-LA	VTLN
1	-	-		
2	-	-		
3	-			
4		-		
5		-		
6				-
7		-		
8			-	-
9		-		
10			-	-
11		-		
12		-		
13			-	
14		-		
15	-			
16		-		
17			-	

A continuación, presentamos las ventajas y desventajas de los algoritmos utilizados para el reconocimiento de la voz.

5. Ventajas e inconvenientes de los algoritmos del reconocimiento de la voz

Una de las ventajas del reconocimiento de la voz en el aprendizaje es ayudar a identificar la pronunciación de un usuario. Algunas de estas aplicaciones como: la corrección de faltas ortográficas [10], para identificar la edad de una persona [15], identificar el género [8].

Para el aprendizaje de los niños mediante dispositivos tecnológicos adaptados a sus necesidades, en especial para niños con problemas de aprendizaje, por ejemplo: los niños con la aplicación de reconocimiento podrá reconocer imágenes, reproducir la voz en idioma extranjera, también podrá reconocer la voz mediante juego, culminando la actividad, la aplicación dará un informe de cómo va el niño con su pronunciación, detallando el nivel de aprendizaje obtenido, además del nivel de vocalización que va logrando el con el sistema de reconocimiento [9].

Por otra parte, es importante citar los inconvenientes que se presentan en el uso del reconocimiento de la voz, como el más destacado está el ruido, que conforma una de las causas principales de la distorsión de la voz. Por otro lado, la fluidez en el habla

del usuario al momento de que el sistema lo identifica, esto sucede con mayor frecuencia en niños, debido en gran parte al temor de hablar con una máquina, cuya causa está relacionada a las emociones, siendo la principal el nerviosismo que causa tartamudez y provoca que la voz no salga de una forma natural y fluida. [1]. Cabe indicar que una de las soluciones para mejorar los problemas con el ruido es el método VTLN, el cual calcula los factores de distorsión de frecuencia para ayudar a tener un mejor reconocimiento.

6. Usuarios finales

El reconocimiento de la voz ha sido utilizado como herramienta de aprendizaje con diferentes fines, desde niños, jóvenes y adultos. Por ejemplo, para corregir faltas ortográficas [2], mediante juegos que consisten en contar cuentos e interactuar con la narración por medio de un robot que identifica la voz [8], aprendizaje de idiomas extranjera [2], redacción de documentos [12], entre otros,

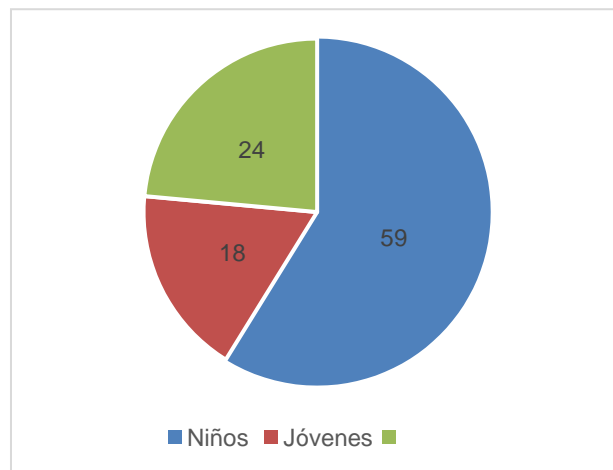


Fig. 2 Usuarios finales

Se puede observar que el reconocimiento de la voz es más utilizado en niños, seguido por jóvenes que interactúan con más facilidad con los dispositivos aprendiendo idiomas o a través de conversaciones, y en adultos mediante la readaptación de un documento o hasta saber la calificación de pronunciación. Otra de las aplicaciones es su apoyo en niños con discapacidad como son los niños no videntes que gracias a esta tecnología podrán redactar documentos mediante el uso del habla [15]. .

7. Resultados

El algoritmo DNN aumentan en gran medida la calidad de analizar la señal de la voz en tiempo real, utilizando reconocimiento profundo y lo complementa con el método

HMM, ya que estos realizan un mejor proceso de reconocimiento de la voz. Para calcular la distorsión de frecuencia, se utiliza el método VTLN [12]. El reconocimiento de la voz es utilizado con mayor porcentaje de nivel escolar en un 59%, seguido de los adultos en un 24%, y un 18% en jóvenes. Entre los algoritmos y métodos más utilizados esta DNN y HMM. Las aplicaciones revisadas, sirven de apoyo y brindan una ventaja a esta tecnología que no solo ayuda a los niños sino como podemos observar es utilizado sin distinción de edad.

8. Conclusiones y recomendaciones

En esta investigación se logró analizar los diferentes tipos de algoritmos y métodos utilizados en el reconocimiento de la voz, lo cual ha servido de gran apoyo en un mayor porcentaje para los niños con problemas en el proceso de aprendizaje. Existe variedad de algoritmos que poseen ventajas y desventajas en su utilización, analizar cada una de ellas proporciona información relevante para elegir el más conveniente al igual que sus métodos. Se recomienda que, en uno de los mayores inconvenientes que posee el reconocimiento de la voz como lo es el ruido, además de evitar la interferencia con VTLN, se puede adaptar una habitación con materiales en las paredes que aislen el ruido del exterior.

Referencias

1. Xu, Ziyi; Strake, Maximilian; Tim Fingscheidt; Concatenated Identical DNN (CI-DNN) to Reduce Noise-Type Dependence in DNN-Based Speech Enhancement 2018 , <http://arxiv.org/abs/1810.11217>
2. Tursunov, Anvarjon; Kwon, Soonil; Hee-Suk Pang.; Discriminating Emotions in the Valence Dimension from Speech Using Timbre Features, 2019, DOI:10.3390/app9122470.
3. Cujar, Ricardo ¹ ; Hernández, Giovanni ¹ ; Cadenas Ocultas de Markov y Modelo de Markov de Máxima Entropía: una comparación en POS Tagging con el corpus AnCora, 2019, <https://bibliotecas.ups.edu.ec:2598/scholarly-journals/cadenas-ocultas-de-markov-y-modelo-máxima/docview/2260411444/se-2?accountid=32861>.
4. Chan, William; Jaitly, Navdeep; Le, Quoc; Vinyals, Oriol; Listen, attend and spell: A neural network for large vocabulary conversational speech recognition, 2016, <https://bibliotecas.ups.edu.ec:2598/conference-papers-proceedings/listen-attend-spell-neural-network-large/docview/1790717942/se-2?accountid=32861>.

5. Christoph Sommer ¹, Daniel W Gerlich, Machine learning in cell biology - teaching computers to recognize phenotypes, 2013, DOI: 10.1242/jcs.123604.
6. Žegklitz, Jan; Pošik, Petr , Symbolic Regression Algorithms with Built-in Linear Regression2017,<http://arxiv.org/abs/1701.03641>.
7. Gualdrón, Oscar Eduardo ^[1] ; Duque Suárez, Oscar Manuel ^[1] ; Chacón Rojas, Mauricio Anastasio ^[1] : Diseño de un sistema de reconocimiento de rostros mediante la hibridación de técnicas de reconocimiento de patrones, visión artificial e ia, enfocado a la seguridad e interacción robótica social, 2013. <https://dialnet.unirioja.es/servlet/articulo?codigo=4966234..>
8. Biomedical Engineering Department, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur.: Malaysia. Fuzzy-based discriminative feature representation for children's speech recognition., 2016,<https://doi.org/10.1016/j.dsp.2014.05.004>
9. Jaebok Kim Khiet P.Truong Vanessa Evers: Automatic temporal ranking of children's engagement levels using multi-modal cues, 2017, <https://doi.org/10.1016/j.csl.2017.12.005>
10. Yeung, G., & Alwan, A. (2018). On the difficulties of automatic speech recognition for kindergarten-aged children. Interspeech 2018. <https://doi.org/10.21437/Interspeech.2018-2297>
11. Rohit Sinha,S.Shahnawazuddin: Assessment of pitch-adaptive front-end signal processing for children's speech recognition ,2017, <https://doi.org/10.1016/j.csl.2017.10.007>
12. Jacqueline Kory, Cynthia Breazeal,Storytelling with robots: Learning companions for preschool children's language development,2014,DOI:10.1109/ROMAN.2014.6926325
13. David J.Miller^aNajah F.Ghalyan^{bc}SudeptaMondal^cAsokRay^d, HMM conditional likelihood based change detection with strict delay tolerance.) ,2017, <https://doi.org/10.1016/j.ymsp.2020.107109>

14. Lasse Rouhiainen © Editorial Planeta, S A. Inteligencia artificial, 2018.
15. Shuaiju Jin a , Xiufeng Wang c , Leilei Du a , Dan He ,Evaluation and modeling of automotive transmission whine noise quality based on MFCC and CNN,2020, <https://doi.org/10.1016/j.apacoust.2020.107562>
16. Prashanth Gurunath Shivakumar1 , Alexandros Potamianos2 , Sungbok Lee1 , Shrikanth Narayanan1,Prashanth Gurunath Shivakumar1 , Alexandros Potamianos2;; Improving speech recognition for children using acoustic adaptation and pronunciation modeling, 2014, https://www.isca-speech.org/archive/wocci_2014/papers/wc14_015.pdf.
17. Xinglei Zhu Gerald T. Beauregard Lonce Wyse;;Real-time iterative spectrum inversion with look-ahead ,2006, DOI:10.1109/ICME.2006.262424.
18. Teng Haikun* Computer and Information Engineering College, Heihe University Academic Road No 1. Speech recognition model based on deep learning and application in pronunciation quality evaluation system ,2019, DOI: 10.1145/3335656.3335657.
19. Feisal Dani Rahman, Noraini Mohamed, Mumtaz Begum Mustafa, Siti Salwah Salim;; Automatic speech recognition system for Malay speaking children ,2014, **DOI:** 10.1109/ICT-ISPC.2014.6923222
20. Hernán Andrade , Lizbeth Pérez , Marlene Sandi , Lizbeth Vildoso y Gabriela Orellana;; Programa psicopedagógico de detección, asistencia y estimulación lingüística de problemas de dislalia en niños de jardines infantiles de la Ciudad de Sucre.2014,,<https://www.ecorfan.org/bolivia/handbooks/ciencias%20sociales%20I/Articulo%2019.pdf>