

<https://doi.org/10.25213/2216-1872.21>



Reconocimiento Digital de Señales de Voz en Colombia

Digital Recognition of Voice Signs in Colombia

Fernando Vitola-De La Rosa

Magister en Entornos Virtuales de Aprendizaje
fernando.vitola@unitecnar.edu.co, <https://orcid.org/0000-0001-6813-8047>,
Docente Fundación Universitaria Antonio de Arévalo - Unitecnar.
Cartagena de Indias, Colombia

*Autor de correspondencia: fernando.vitola@tecnar.edu.co



© 2019 Fundación Universitaria Antonio de Arévalo – UNITECNAR.

Cómo citar: F. Vitola De La Rosa, "Reconocimiento Digital de Señales de Voz en Colombia", *Sostenibilidad, Tecnología y Humanismo*, vol. 10, no. 2, 65-73, 2019.

Resumen

El reconocimiento de voz y los avances tecnológicos sean hecho cada vez más importantes en nuestra era, una fascinación por entender el comportamiento de las mismas se hacen presente, siempre ha sido un hecho curioso el poder comprender como proceso exacto, a través de que se puede medir o distinguir la voz, la implicación de un ordenador es la clave fundamental, así habría un entendimiento de hombre-máquina, por supuesto mediante una digitalización en código binarios que el ordenador pueda comprender, más allá de que sea curiosidad el mismo se realiza con el fin de resolución de problemas sobre todo en aquellos casos que personas puedan tener alguna discapacidad en el modulado de sus palabras, y traducir, o en su defecto la búsqueda rápida de palabras en la internet,; por lo tanto se hace fascinate conocer las diferentes tecnología o programas que permiten la síntesis del reconocimiento de voz, por consiguiente, el presente este artículo se realiza un planteamiento general del problema del procesamiento de voz, se hace un análisis de los principales avances que se han presentado a nivel mundial en reconocimiento, síntesis y codificación de voz al igual que algunos de los trabajos realizados en Colombia en esta área.

Palabras clave: Codificación, procesamiento, reconocimiento, síntesis, voz

Abstract

Voice recognition and technological advances are becoming increasingly important in our era, a fascination to understand their behavior is present, it has always been a curious fact to understand as an exact process, through which you can measure or distinguish the voice, the implication of a computer is the fundamental key, so there would be an understanding of man- machine, of course by means of a digitalization in binary code that the computer can understand, beyond curiosity it is done with the purpose of solving problems especially in those cases where people may have some disability in the modulation of their words, and translate, or failing that, the rapid search for words on the internet; therefore it becomes fascinating to know the different technology or programs that allow the synthesis of voice recognition, therefore, this article presents a general approach to the problem of voice processing, an analysis of the main advances that are made have presented worldwide recognition, synthesis and voice coding as well as some of the work done in Colombia in this area.

Keywords: Coding, processing, recognition, synthesis, voice

Introducción

Hoy día la globalización tecnológica va de la mano con la evolución del mundo, el cual avanza vertiginosamente con el afán de seguir buscando en innovando el procesos diarios, por lo tanto, las tecnología de información y comunicación pasan a ser brazo ejecutor para la resolución de problemas antes nuevos retos; así lo refleja [1] en “un mundo altamente complejo como el actual expresa una profunda crisis”, crisis que conlleva la búsqueda de nuevas herramientas tecnológicas que permita la comunicación los seres humanos.

De igual manera, [1] indica que “la comunicación efectiva y asertiva se ha hecho cada vez más importante en el día a día”, esto a través de una forma natural como lo es el habla, generando con el transcurrir del tiempo diferentes sistemas que emulen este tipo de acciones. Lo complejo de este proceso se ha abordado desde diferentes puntos de vista, dando lugar a varias áreas de investigación tales como: Reconocimiento de voz, reconocimiento de hablante, síntesis de voz, codificación de voz; así lo refleja [2] “desde sus inicios, el hombre ha sentido la necesidad de relacionarse y comunicarse con el mundo que lo rodea, razón por la cual a través de los años hemos estado en constante búsqueda de aquellos métodos que nos ayuden a suplir dicha necesidad”.

De igual manera, la búsqueda constante de solucionar problemas del habla y la escucha siempre ha estado presente, sobre todo para personas con dificultades auditivas, así lo refleja [2] “en 1940 para ser exactos, se inicia el desarrollo de una tecnología capaz de crear un enlace. Entre personas y máquinas conocida como “reconocimiento de voz”.

La misma surge a raíz de querer construir un sistema que hiciera el habla visible a las personas que presentaban problemas auditivos”, por lo tanto, el desarrollo de sistemas

computarizados orientados a la adquisición, el análisis, el procesamiento y la reproducción de la voz, en los que se emplean modernas técnicas de procesamiento digital de señales ha tenido un importante auge durante las últimas décadas, con diversas aplicaciones lingüística, fonética, psicolingüística y análisis de la voz normal [3].

Según [4] “un sistema de reconocimiento de voz no es más que la capacidad que posee un ordenador, de convertir, las palabras de la voz humana a un código binario comprensible por computador”.

Las palabras reconocidas pueden ser el resultado final en algunas aplicaciones, mientras que, en otras, éstas se utilizan como entradas para realizar un procesamiento adicional que conlleve al entendimiento del mensaje, siempre de la mano con las nuevas tecnologías de información y comunicación, a través de los nuevos algoritmos que permita la digitalización de la voz en códigos binarios. “Quienes apuesten por estas herramientas estarán en una mejor posición para lograr un lugar privilegiado en el futuro que hoy ya se empieza a diseñar” [5].

Así mismo, [6] expresa “la necesidad de detectar y determinar cuáles serían los elementos en este caso el reconocimiento de voz que distinguen los sistemas que trabajan con palabras aisladas, en los cuales las palabras deben pronunciarse con grandes pausas entre ellas, de los sistemas de voz continua en donde las palabras se pronuncian de manera natural. También es necesario especificar si el sistema es dependiente o independiente del hablante, si se trabaja con vocabularios reducidos (menos de 20 palabras) o con grandes vocabularios (más de 20000 palabras).



Figura 1. Señal de voz.

El reconocimiento de voz no es una tarea fácil, debido a las fuentes de variabilidad asociadas con la señal de voz, así lo sostiene [7] “algunos de estos problemas son resueltos por los humanos sin mucho esfuerzo. Sin embargo, en muchos casos, la solución a estos problemas, usando computadoras, se vuelve extremadamente difícil” Por ejemplo, los fonemas son muy dependientes del contexto, y suenan diferente en las distintas palabras. Por otra parte, puede haber variaciones acústicas debidas a cambios en el ruido de fondo o en la posición y tipo de transductor utilizado. Otras variaciones de la señal de voz se deben al hablante, ya que incluso si se trata de la misma persona, ésta puede pronunciar las palabras de manera diferente dependiendo de su estado de ánimo.

Finalmente, otro o incluso en las diferentes regiones de un país, se pueden tener diferentes formas de hablar en el mismo idioma.

Campos de investigación sobre procesamiento de señales de voz.

En 1992 la Asociación Nacional de Ciencia de los Estados Unidos indicó que los campos hacia los cuales deberían orientarse las investigaciones relacionadas con el reconocimiento de voz eran las siguientes:

Robustez, Portabilidad, Adaptación, Modelado del lenguaje, Medida de la confiabilidad, Palabras fuera del vocabulario, Habla espontánea, Modelado dinámico y

Prosodia; estas variables aportando para una investigación más profunda del proceso de la digitalización del reconocimiento de voz a través de un ordenador; una realidad latente es que estamos arropado por una era de inteligencia artificial, influyendo en la generación de voz artificial. La mayoría de las técnicas de síntesis de voz se encargan de modelar mediante filtros las características del aparato fonador humano, y a partir de este modelo se reproducen los diferentes sonidos que conforman la voz. En este campo se tiende a buscar que la voz generada sea lo más natural posible y una de las principales aplicaciones es la conversión de texto a voz y la comunicación máquina y ser humano. Así lo refleja [8] “un sistema de reconocimiento de voz es el encargado de establecer la comunicación entre los humanos y las computadoras con el desarrollo de aplicaciones capaces de reconocer diferentes voces en la medida de que el habla sea de manera natural.”



Figura 2. Robot hablante.

Otra tarea muy común en el procesamiento de señales de voz es el reconocimiento del hablante, en la cual el sistema tiene la capacidad de identificar quien es la persona que habla dentro de una población previamente establecida. Debe diferenciarse el reconocimiento de la verificación del hablante, donde se determina por medio de la voz si

la identidad suministrada por el usuario es correcta o si se trata de una persona diferente. En estos casos no tiene mucha importancia el reconocimiento del mensaje. El procesado digital de voz ha sido una de las áreas de mayor trabajo en el campo del procesamiento digital de señales. Para entender cómo reconocer los patrones de una señal de voz, se debe primero entender el funcionamiento del aparato fonador humano [9].

Por otra parte, la codificación de voz consiste en almacenar de forma eficiente la señal de voz, de tal manera que a partir de los parámetros codificados pueda volverse a generar la señal original. Este campo ha sido bastante estudiado, tratando de diseñar codificadores que permitan obtener voz de buena calidad y consumiendo la mínima cantidad de recursos. Como resultado de las investigaciones en codificación de voz se han desarrollado muchos códecs, algunos se han convertido en estándares y otros son sistemas propietarios para aplicaciones especiales. Estos codificadores de voz han permitido transmitir la voz sobre canales de comunicación y redes de paquetes como Internet. El proceso de reconocimiento de voz inicia con la conversión de una señal de voz a una secuencia de palabras, por medio de un algoritmo implementado como un programa de ordenador [10].

Evolución del procesamiento de señales de voz

Según [11] sostiene que “uno de los campos más interesantes de la inteligencia artificial es la creación de interfaces entre el hombre y la máquina cada vez más amigables, una de las vertientes de este campo es el reconocimiento de voz, cosa que por ser de carácter simbólico es altamente compleja desde el punto de vista computacional”, los mismo se evidencia en los paneles de la historia de la computación publicados por IEEE en el año 2006, se describen los principales avances

en el procesamiento de voz obtenidos por BBN, uno de los principales laboratorios a nivel mundial en el área del procesamiento de señales de voz. A continuación, se resumen estos avances para acercarnos al estado del arte del procesamiento de señales de voz a nivel mundial:

Hasta el año de 1970 la gran mayoría de los trabajos en este campo estaban relacionados con el reconocimiento de palabras aisladas y con diccionarios de tamaño reducido que, en la mayoría de los casos trabajaban con sistemas analógicos debido a la gran dificultad que se presenta en el proceso de reconocimiento de voz, así como lo refleja [12] “Un reconocimiento de voz tiene la dificultad principal en la extracción de características de la voz, debido a que cuenta con diversos problemas a solucionar”.

En el período 1972 – 1991 los trabajos se orientaron a la codificación y compresión de la voz. La operación pretendía que se pudiera transmitir voz sobre una red de datos lográndose reducir la tasa de transmisión de voz de 64 Kbps hasta 2400 bps con una calidad de voz razonable. La primera demostración de comunicaciones bidireccionales de voz utilizando voz comprimida sobre una red de paquetes se realizó en 1974. A partir de este proyecto se iniciaron otros similares en los cuales se obtuvieron tasas de 2.4, 9.6 y 16 kbps tratando de mejorar la calidad de la señal de voz recibida. También se emplearon técnicas avanzadas de cubanización vectorial que permitieron realizar comunicaciones de voz a tasas muy bajas, entre 200 y 300 bps. De tal manera que [8] sostiene que “la comunicación digital es una expresión general que puede cubrir cualquier intercambio de información codificada. En un sentido más restringido puede definirse como la “transmisión de datos entre dos o más dispositivos terminales”. Esta definición restringida hace aparecer el importante concepto de “interfaz” entre la

fuelle generadora de datos y el canal digital propiamente dicho.

Posteriormente entre 1982 y 1988 se realizaron trabajos enfocados hacia la modificación de las características de la voz para diferentes aplicaciones y el mejoramiento de las características de la voz sintética con el objetivo de que ésta sonara mucho más natural. Para esta misma época se desarrolla un proyecto cuyo objetivo era el reconocimiento del hablante, realizando las tareas de identificación y verificación. Un nuevo aporte en el área del reconocimiento del habla realizado por este proyecto fue la utilización de los Modelos Ocultos de Markov (HMM), los cuales mejoraron ampliamente los resultados que se habían obtenido hasta el momento y cambiaron el paradigma del reconocimiento de voz, lo cual condujo a ARPA a iniciar un nuevo proyecto de reconocimiento basado en HMM, que pudo ser extendido a otros idiomas sin necesidad de escribir nuevas reglas. Para el año de 1987 el sistema desarrollado efectuaba el reconocimiento continuo de un vocabulario de 1000 palabras casi en tiempo real, y en 1990 se obtuvo el reconocimiento en tiempo real. Las señales de voz están formadas por secuencias de sonidos. Estos sonidos y la transición entre ellos llevan la información que necesita ser transportada [13].

La primera demostración de un reconocedor de voz continua de 20000 palabras en tiempo real se realizó en el MIT en 1993, lo cual se logró gracias al aumento de la velocidad y capacidad de memoria de los procesadores y al desarrollo de algoritmos de procesamiento de voz mucho más eficientes. En este punto es importante destacar que en 1987 se realizaron trabajos denominados “Word trabajos más recientes desarrollaron sistemas ópticos de reconocimiento de caracteres (OCR) para diferentes idiomas. Así lo sostiene [12] “los sistemas de reconocimiento de voz con el tiempo han tenido un gran auge en la sociedad

moderna, ante la creciente necesidad de tener sistemas que se puedan controlar de manera no física, para un porcentaje creciente de personas discapacitadas en el mundo”.

Trabajos sobre procesamiento de señales de voz en Colombia.

En Colombia, el procesamiento de señales de voz en sus distintas facetas se ha convertido en uno de los temas más llamativos para tesis de grado, proyectos de grupos de investigación, ponencias en congresos, productos tal vez de la gran cantidad de discapacitados que cada día entra a la estadística de los mismo, de acuerdo a las últimas estadísticas sobre discapacidad publicadas en Colombia por el DANE [13], de un total de 84.283 personas censadas, 1.036 respondieron que tenían deficiencias. A continuación, se mencionan algunos de los trabajos realizados en los últimos años en el ámbito académico colombiano:

- Experimentación en reconocimiento de voz -dígitos aislados- mediante cuantización de vectores de predicción lineal y modelos escondidos de Markov se presentan dando como resultado porcentajes superiores al 90% para condiciones normales del ruido de fondo [15-16].
- Verificación del hablante utilizando técnicas de modelos ocultos de Markov [17].
- En el festival de Linux organizado por la Universidad Tecnológica de Pereira, el Grupo Linux del Eje Cafetero presentó la ponencia “Tecnologías de Reconocimiento de voz y su Implementación en Software Libre, una aplicación alternativa para personas discapacitadas”.
- El grupo de investigación en Bioingeniería de la Universidad Javeriana, ha desarrollado varios proyectos relacionados con la codificación visual del habla, algunos de los

cuales han sido financiados por Conciencias, los cuales tienen como objetivo la transmisión de telefonía visual por canales de comunicación de muy baja velocidad [18].

- La Universidad Nacional de Colombia, Sede Manizales ha trabajado fuertemente en las diferentes áreas del procesamiento de voz, tales como reconocimiento, Verificación del hablante, codificación, voces patológicas [19] entre otros [20-22].

Conclusiones

A manera de conclusión, puede observarse que el procesamiento de voz ha sido un área bastante trabajada en nuestro país, y basta con revisar las tesis de grado o los productos de investigación de las principales universidades colombianas para encontrar trabajos relacionados con este tema. Aportando a la creación de algoritmos para el procesamiento del mismo, aun cuando se encuentran deficiencias en el proceso digital del reconocimiento de voz, sigue la búsqueda en la mejora mediante el uso de la inteligencia artificial y las nuevas tecnologías de información y comunicación. Aún queda mucho camino por recorrer en este campo de estudio, por lo tanto se hace necesario motivar a los estudiantes de ingeniería desarrollar proyectos para este objeto de estudio, en el este artículo se logró evidenciar el gran porte social que tiene el estudio de la digitalización de reconocimiento de voz, enfocados al gran número de discapacitados auditivos a nivel territorial, siendo esta herramienta un traductor que abre un espacio para los que no pueden ser escuchados o en su defecto aportar en la escucha de los que no pueden oír.

Conclusiones

A manera de conclusión, puede observarse que el procesamiento de voz ha sido un área bastante trabajada en nuestro país,

y basta con revisar las tesis de grado o los productos de investigación de las principales universidades colombianas para encontrar trabajos relacionados con este tema. Aportando a la creación de algoritmos para el procesamiento del mismo, aun cuando se encuentran deficiencias en el proceso digital del reconocimiento de voz, sigue la búsqueda en la mejora mediante el uso de la inteligencia artificial y las nuevas tecnologías de información y comunicación. Aún queda mucho camino por recorrer en este campo de estudio, por lo tanto se hace necesario motivar a los estudiantes de ingeniería desarrollar proyectos para este objeto de estudio, en el este artículo se logró evidenciar el gran porte social que tiene el estudio de la digitalización de reconocimiento de voz, enfocados al gran número de discapacitados auditivos a nivel territorial, siendo esta herramienta un traductor que abre un espacio para los que no pueden ser escuchados o en su defecto aportar en la escucha de los que no pueden oír.

Referencias

- [1] J.J. Rueda-López, “La tecnología en la sociedad del siglo XXI: Albores de una nueva revolución industrial”, *Aposta. Revista de Ciencias Sociales*, no. 32, pp. 1-28, enero-marzo, 2007
- [2] K. Barrios, J. López, S. Mendieta, R. Benavides y Y. Sáez, “Sistema de reconocimiento de voz: un enlace en la comunicación hombre-máquina”, *Revista de iniciación científica*, vol. 4, pp. 92-95, 2018
- [3] Z. Furui, *Digital Speech Processing – synthesis and recognition*. USA, 1989
- [4] M. Esparza, M. Avalos y J. Benito, “Reconocimiento de voz”, *Conciencia Tecnológica*, no. 22, 2003.
- [5] J. Spector, “Emerging educational technologies and research directions”, *Journal of educational*

technology & society, vol. 16, no. 2, 2013.

- [6] Z. Alezones-Campos y Y. Baquero-Romero, “Reconocimiento de palabras aisladas para control de navegación de robot móvil”, *Ciencias Básica e Ingeniería*, vol. 16, no. 1, 2012
- [7] C.M Bishop, *Neural Network for Pattern Recognition*, USA: Oxford University Press, 1995
- [8] L. Ruffiner, D. Milone, “Sistema de reconocimiento automático del habla”. *Ciencia, Docencia y Tecnología*, vol. XV, no. 28, pp. 151-177, 2004
- [9] J. Gómez, J. Simancas, M. Acosta, F. Meléndez y J. Vélez, “Algoritmo de reconocimiento de comandos voz basado en técnicas no-lineales”, *Espacios*, Vol. 38, No. 17, 2017
- [10] M Tahir, and S. Ashfaque, “Voice Controlled Wheelchair Using DSK TMS320C6711”, in *International Conference on Signal Acquisition and Processing*, IEEE, 2009
- [11] J. Estard, “Algunas experiencias sobre reconocimiento de fonemas utilizando redes neuronales artificiales.” *Ingeniería Electrónica, Automática y Comunicaciones*, vol. 21, no. 2, 2000
- [12] C. De Luna, J. Martínez, M. Mora, “Reconocimiento de Voz con Redes Neuronales, DTW y Modelos Ocultos de Markov”, *Conciencia Tecnológica*, no. 32, 2006
- [13] M. Latinus and P. Belin, “Human voice perception”, *Current Biology*, vol. 21, no. 4, pp.143–145, 2011
- [14] Departamento Administrativo Nacional de Estadística, *Información Estadística de la Discapacidad*, 2004. [En Línea]. Disponible en: http://www.dane.gov.co/files/investigaciones/discapacidad/inform_estad.pdf.
- [15] C. Rueda, “Experimentación en reconocimiento de voz -dígitos aislados- mediante cuantización de vectores de predicción lineal y modelos escondidos de Markov”, Proyecto de grado. Facultad de Ingeniería de Sistemas. Universidad Autónoma de Manizales. Colombia. 1993
- [16] C. Rueda, “Aplicación de la cuantización vectorial en la verificación automática del hablante”, *Anfora*, Vol. 5, No. 9, pp. 44-58, 1997
- [17] F. Silva F. Verificación del hablante con cadenas escondidas de Markov. Universidad Nacional Sede Manizales. 2004
- [18] J. Villegas, C. Soto, L.R Valderrama, R. Solano et al., “Codificación de Telefonía Visual Empleando Conjunto de Imágenes”. *Revista Energía y computación*. Vol. XIII, no. 2, pp.36 - 42 ,2005
- [19] C. Ojeda, “Extracción de características utilizando transformada Wavelet en la identificación de voces patológicas”. Tesis, Universidad Nacional Sede Manizales, 2003
- [20] F. Betancourth y F. Vitola, “Reconocimiento en tiempo real de palabras aisladas de un vocabulario reducido en Idioma español”. Tesis, Universidad Nacional de Colombia. Manizales. 1999
- [21] L. Romo, “Síntesis de voz aplicada a la rehabilitación de pacientes con labio y paladar hendido corregido”. Trabajo de grado, Universidad Nacional Sede Manizales. 2005
- [22] F. Sepúlveda, “Extracción de parámetros de señales de voz utilizando técnicas de tiempo y frecuencia”. Tesis, Universidad Nacional Sede Manizales. 2004