# Antecedentes y Fonetistas de Tecnologías del Habla

**Alma Delia Sánchez Rivero**
Universidad Autónoma del Carmen
**almasanchezrivero@hotmail.com**
**asanchez@pampano.unacar.mx**

## Resumen

***Antecedentes y Fonetistas de Tecnologías del Habla*** es producto del trabajo de investigación bibliográfica sobre Tecnologías del Habla (TH). En este trabajo se conjunta el análisis de artículos en diferentes lenguas en la temática de la trascendencia del descubrimiento que marcan el rumbo de la ciencia y la tecnología del habla así como del papel del fonetista en un grupo de TH y un ejemplo de aplicación de las TH en el campo de la jurisprudencia.

## Abstract

The background and all phoneticians of Speech Technologies is a product of the work of bibliographic research on Speech Technologies (TH). In this work the joint analysis of articles in different languages in the thematic of the significance of the discovery that set the course for science and technology of speech as well as the role of the fonetista in a group of TH and an example of application of the TH in the field of jurisprudence

Palabras clave/ Key Words: fonetistas, tecnologías del habla, investigación bibliográfica, diferentes lenguas, ciencia, papel del fonetista, TH, jurisprudencia/ Phoneticians, speech technologies, bibliographic research, different languages, science, the role of the fonetista, TH, jurisprudence.

---

## Introducción

En **antecedentes** se describen hechos trascendentes, tanto resultados de investigación científica como publicaciones que han revolucionado la tecnología del habla. **Los fonetistas en la tecnología** del habla describen el desempeño del trabajo del fonetista dentro de un grupo de desarrollo de las tecnologías del habla en las carreras de lingüística y fonética.

En **aplicaciones** se da un ejemplo de la utilización de la ciencia y la tecnología del habla en la que un grupo de expertos da una opinión de la posibilidad o no de identificar con certeza a una persona por su voz con implicaciones legales y judiciales.

**Antecedentes**

En febrero de 1989 Lawrence Rabiner publicó *A Tutorial on Hidden Markov Molds* and *Selected Applications in Speech Recognition*; es un artículo científico que al día de hoy es un paradigma en el origen de los métodos estadísticos y modelaje de los procesos de reconocimiento.

Estos modelos ocultos de Markov son muy ricos en estructura matemática y forman la herramienta básica teórica en este tipo de modelaje estadístico que bien aplicado trabaja aplicaciones importantes en máquinas de reconocimiento del habla.

Posteriormente en 1993 Lawrence Rabiner junto con Biing-Hwag Juang publicaron "Fundamentals of Speech Recognition" un libro completo con las bases de reconocimiento del habla y cuyo objetivo es "proporcionar un sonido teórico y técnicamente exacto y una descripción completa y razonable del comportamiento básico e ideas que constituyen un sistema moderno de reconocimiento del habla por máquina"[1]; demuestra también que "aunque existe una base sólida para la descripción lingüística del sonido, y con un buen entendimiento de la acústica en la producción de sonido existe en el mejor de los casos, una tenue relación entre un sonido lingüístico dado y un repetible, seguro y medible conjunto de

[1]Rabiner, L. y Biing-Hwag Juang, 1993:xxxi.

parámetros acústicos".[2] También prueba que el nivel básico del habla o el reconocimiento del sonido es sólo un peldaño en un largo proceso donde tareas de alto nivel de información en la forma de sintaxis, de semántica y pragmatismo a menudo juegan un mejor papel.

En el desarrollo histórico de los mejores avances de la tecnología del habla (Myers, Brad, 1996) se enfatiza la importancia de la investigación científica en las universidades y la interacción hombre máquina. (HIC).

Las interacciones básicas como la manipulación directa de objetos gráficos donde objetos visibles en la pantalla son manipulados directamente con un objeto puntual fue demostrado por primera vez por Iván Sutherland en 1963 con su tesis doctoral en el MIT, el desarrollo del Mouse en el laboratorio de investigación de Stanford en 1965 y Windows en 1968 son ejemplos en el campo de la HIC del trabajo científico universitario.

**Los fonetistas en la tecnología del habla**

La tecnología del habla es una excitante carrera con futuro promisorio, comenta Alejandro Acero (1995) por la alta inversión financiera de las compañías en conjunto con universidades alentadas por los avances en tecnología computacional y los estudios de investigación de mercados.

Se hace mención de la existencia de paquetes de programación que realizan muy limitadamente tanto reconocimiento automático por voz RAV (ASR abreviados en inglés) como sistemas Texto al habla TAH (TTS abreviado en inglés).

La posición de los fonetistas en la tecnología del habla es una descripción del desempeño del trabajo multidisciplinario que compartirán con ingenieros y científicos de computación.

Existe una distinción de objetivos entre las ciencias y las tecnologías del habla en la que en las ciencias se busca el conocimiento de mecanismos de producción del habla y en las tecnologías el de construir sistemas de lenguaje hablado.

Definiendo un sistema de lenguaje hablado como una circuitería corriendo un programa de cómputo o de ordenador, enlista las tareas de construir tablas de datos primarios y secundarios, reglas y algoritmos para un lingüista o fonetista en un equipo de tecnologías del habla.

[2]Rabiner, L. y Biing-Hwag Juang, 1993: xxxii.

En la tabla de datos primarios describe los componentes y establece la importancia del diseño cuidadoso de componentes como *el alfabeto de fonemas para una lengua, el mapeador entre el diccionario y el alfabeto de fonemas, el mapeador entre símbolos y palabras del diccionario y el modelador del lenguaje.*
Las tablas son resultado de procesar algún corpus con un programa de cómputo y la tabla de datos secundarios tiene como tarea obtener estos corpus etiquetados como por ejemplo listas de oraciones balanceadas fonéticamente, colección de datos para TAH, anotación de datos para TAH, corrección de datos para RAV y anotación de datos para RAV.

La mejoría de **las reglas y algoritmos** implica una mejor realización del sistema y un fonetista, describe Alejandro Acero en su ensayo, por ejemplo puede participar "surgiendo el mapeo entre todos los posibles contornos de entonación que pueden ocurrir en un lenguaje y sus dependencias con el tipo de oración".
Desde este punto de vista laboral concluye con los requerimientos del empleo de la posición de los fonetistas en las tecnologías del habla y se hace hincapié de la importancia del entretenimiento y el trabajo de un equipo en el que "cada miembro tiene que tener un entendimiento básico del sistema".
Establece Alejandro Acero que además de los cursos normales de lingüística un fonetista trabajando en un equipo de tecnologías del habla debería tomar un curso de fundamentos de computación y programación básica. También comenta que todos los miembros del equipo pudieran hablar el mismo lenguaje técnico científico, así la importancia de los avances en el procesamiento de señales y estadística, de los entornos de cómputo como Windows, Unix, editores de texto, y paquetes de análisis de habla.
Concluye que el objetivo principal para un sistema RAV es mejorar el conocimiento con precisión y en el sistema TAH es mejorar la calidad del habla, no perder la dirección de estas líneas y que cada miembro del equipo no pierda el enfoque global.
Este trabajo sólo reporta una referencia de Meisel W.S. (1992).

**Aplicaciones**

**Identificación de personas por la voz**

En la actualidad no hay procesos científicos que permitan caracterizar la voz de una persona o identificar con absoluta certeza a un individuo por su voz.

El artículo de Bonastre ***Authentification des personnes par leur voix: un nécessaire devoir de précaution*** está destinado a todo tipo de público. Discute los límites científicos y tecnológicos de las técnicas y métodos de la identificación de los individuos por medio de la voz (autentificación vocal). Los investigadores en el procesamiento del habla exponen su punto de vista sobre la fonética en el contexto de la criminalística.

Los métodos para la identificación de voz no siempre están apoyados en un planteamiento científico. En el estado actual de conocimientos, no existen procedimientos, automáticos o manuales, que permitan afirmar con certeza que una persona es –o no es- la autora de un registro vocal dado. Esto es más verdadero en cuanto se trata de autentificar un registro de duración limitada, con ruido de fondo, registrado en malas condiciones técnicas y procediendo de una persona que disfraza o modifica su voz.

Existe la capacidad natural del ser humano de reconocer a un orador, pero esta capacidad está influida por varios factores. La familiaridad, la duración de los ejemplos sonoros, el contexto, el intervalo temporal entre los ejemplos, las condiciones de tensión y modificación voluntaria de la voz, el control de los auditores.

El espectrograma es una herramienta útil para el tratamiento y análisis de la voz. Una “empreinte vocable” es simplemente un espectrograma de una señal de vocal que puede estar impresa. Se trata de un gráfico que representa la señal en tres dimensiones: el tiempo, la frecuencia y la intensidad.

Que sea “vocal impresa” no significa que tenga el mismo nivel de unicidad y fiabilidad que las huellas genéticas. Las investigaciones científicas no permiten afirmar que la voz posee características que logren identificar de una única manera al ser humano.

La voz presenta diferencias importantes con las huellas genéticas. La voz evoluciona durante el tiempo, a corto plazo (durante el día), a medio plazo (al año) y a largo plazo (con la edad), así que está en función del estado de salud o emocional. La voz es modificable voluntariamente, existen técnicas. En el campo de la

vocal, las bases de datos disponibles no implican un número suficiente de oradores, de lenguas y de condiciones registradas para la evaluación de métodos de autentificación, a alto nivel de fiabilidad.

En el juicio Daubert V. Merrell Dow Pharmaceuticals rendido en 1993, el tribunal de los Estados Unidos decidió que 5 condiciones debían cumplirse para que un elemento de prueba pueda considerarse como científico en un tribunal: el método debe probarse o poder probarse, el método se ha sometido a crítica de los pares, existen normas puestas al día que controlan el uso de la técnica, la técnica es aceptada por la comunidad científica y el potencial de errores debe conocerse (y ser aceptados).

Habida cuenta de los factores de Daubert se plantean distintas cuestiones respecto al método espectrográfico: ¿es que una prueba basada en espectrogramas puede ser considerada científica? ¿La objetividad de la comparación de las señales está garantizada? ¿cuál es el nivel conocido potencial de errores? ¿cuáles porcentajes de errores provienen respectivamente del método y de su mismo análisis? ¿qué comunidad científica reconoce la técnica espectrográfica?

Las técnicas de reconocimiento de voz están basadas en medidas de semejanza de los registros de palabras. Estas medidas están hechas sobre parámetros acústicos extraídos por análisis de señal. Se puede tener en cuenta información específica del orador, el contenido del mensaje vocal, la información sobre el medio ambiente y el material de registro.

Para garantizar un nivel de resultado aceptable para las aplicaciones del reconocimiento de voz, son generalmente necesarias otras características: los oradores no deben intentar disfrazar la voz, las condiciones de registro y de tratamiento de señal de audio son conocidas y/o controladas. Los datos de palabra, registrados en las mismas condiciones que la señal de prueba, deben estar disponibles para hacer referencia a un orador en el sistema. La medida de semejanza se calibra durante experiencias realizadas en las condiciones controladas citadas anteriormente. El método de decisión se estima en función de resultados de experiencia y en función de la aplicación contemplada.

Asimismo, añadir dificultades pueden dar mejores resultados: el uso de técnicas sofisticadas para modificar o disfrazar la voz debe estar prohibido a los impostores potenciales. No se autoriza el uso de un sistema de síntesis de palabra. El contenido lingüístico de mensajes incluye palabras conocidas del sistema, permitiendo a éste calcular una semejanza entre voces basándose en contenidos comparables.

Las investigaciones actuales están concentradas en las limitaciones prácticas del reconocimiento de la voz, la difusión y la interpretación de los resultados del sistema. Los sistemas automáticos pueden ser útiles en adición a otros métodos para ayudar a la orientación de investigaciones cuando los elementos vocales cruciales están disponibles. Las limitaciones mencionadas deben considerarse al momento de la interpretación de los sistemas automáticos.

Precaución y reflexión deben aplicarse en un método de reconocimiento de voz –basada en una competencia humana o automática. Un uso juicioso de estas técnicas puede ser aceptable siempre que no sean dados por infalibles.

## Bibliografía


- Acero, Alejandro. ***The role of phoneticians in Speech technology*** . Microsoft Corporation, Redmond, WA, USA. http://research.microsoft.com/srg/papers/1995-alexac-esca.pdf
- Bonastre, Jean-Francois, FrédericBimbot, Louis-Jean Boë, Joseph P. Campbell, Douglas A. Reynolds e IvánMagrin-Chagnolleau. ***Authentification des personnes par leur voix: un nécessaire devoir de précaution***
- Universitéd'Avignon, France, Universitare de Beaulieu, MIT Lincoln Laboratory, Lexington, Massachusetts USA, DDL, CNRS& University Lyon, Berthelot, Lyon, France. http://www.lia.univ-avignon.fr/fich_art/581-AFCP_SpLC_JEP04_preprint.pdf
- Irigoyen, Maitena y Eloy Irigoyen Gordo ***Reconocimiento del habla***. Universidad del País Vasco. http://www.disa.bi.ehu.es/spanish/asignaturas/17223/Reconocimiento_Habla.pdf
- Myers, Brad A. ***A brief history of human computer interaction technology*** ACM interactions. Vol. 5, no. 2, March, 1998. Pp. 44-54.
- Rabiner, Lawrence R. ***A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition***
- Proceeding of the IEE, Vol. 77, No. 2, Febrero, 1989. Pp. 257-286.
- Rabiner, Lawrence R. y Biing-Hwang Juang . **"Fundamentals of Speech Recognition".**
- Englewoods Cliffs, New Jersey, 1993.