



Ingeniería

ISSN: 0121-750X

revista\_ing@udistrital.edu.co

Universidad Distrital Francisco José de

Caldas

Colombia

Gamma, Erika Nathalia; Amaya Hurtado, Darío; Ramos Sandoval, Olga Lucia

Revisión de las tecnologías y aplicaciones del habla sub-vocal

Ingeniería, vol. 20, núm. 2, 2015, pp. 277-288

Universidad Distrital Francisco José de Caldas

Bogotá, Colombia

Disponible en: <http://www.redalyc.org/articulo.oa?id=498850181008>

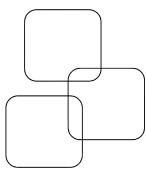
- ▶ Cómo citar el artículo
- ▶ Número completo
- ▶ Más información del artículo
- ▶ Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal  
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

# Artículo de Revisión



Citación: Gamma, E. N., Amaya, D. y Ramos, O. L. (2015). Revisión de las tecnologías y aplicaciones del habla sub-vocal.  
En: Ingeniería, Vol. 20, No. 2, pp. 277–288  
©Los autores. Titular de los derechos de reproducción: Universidad Distrital Francisco José de Caldas. En línea DOI:  
<http://dx.doi.org/10.14483/udistrital.jour.reving.2015.2.a07>

## Revisión de las tecnologías y aplicaciones del habla sub-vocal

### *Technologies and applications review of subvocal speech*

Erika Nathalia Gamma

Universidad Militar Nueva Granada  
gav@unimilitar.edu.co

Darío Amaya Hurtado

Universidad Militar Nueva Granada  
dario.amaya@unimilitar.edu.co

Olga Lucia Ramos Sandoval

Universidad Militar Nueva Granada  
olga.ramos@unimilitar.edu.co



## Resumen

Se presenta una revisión de estado de las principales temáticas aplicativas y metodológicas del habla sub-vocal o lenguaje silencioso que se han venido desarrollando en los últimos años. Entendiendo por habla sub-vocal como la identificación y caracterización de las señales bioeléctricas que llegan al aparato fonador, sin que se genere producción sonora por parte del interlocutor. La primera sección hace una profunda revisión de los métodos de detección del lenguaje silencioso. En la segunda parte se evalúan las tecnologías implementadas en los últimos años, seguido de un análisis en las principales aplicaciones de este tipo de habla; y finalmente se presenta una amplia comparación entre los trabajos que se han hecho en industria y academia utilizando este tipo de desarrollos.

**Palabras claves:** Implantes auditivos, micrófonos, procesamiento del habla, reconocimiento del habla, ruido acústico, síntesis del habla, trabajo en ambientes peligrosos.

## Abstract

This paper presents a review of the main applicative and methodological approaches that have been developed in recent years for sub-vocal speech or silent language. The sub-vocal speech can be defined as the identification and characterization of bioelectric signals that control the vocal tract, when is not produced sound production by the caller. The first section makes a deep review of methods for detecting silent language. In the second part are evaluated the technologies implemented in recent years, followed by a review of the main applications of this type of speech and finally present a broad comparison between jobs that have been developed in industry and academic applications.

**Key words:** Acoustic noise, auditory implants, microphones, speech processing, speech recognition, speech synthesis, working environment noise.

Recibido: 28-04-2015  
Modificado: 19-08-2015  
Aceptado: 09-09-2015

# 1. Introducción

El lenguaje silencioso o habla sub-vocal está relacionado con la extracción, caracterización y clasificación de señales biológicas que provienen de la corteza cerebral, viajan por el sistema nervioso central, hasta llegar al aparato fonador. Estas señales biológicas representan la intención del habla antes de ser articulada en voz alta. Las interfaces de habla silenciosa (SSI-Silent Speech Interfaces, por sus siglas en inglés), se definen como el proceso de adquisición y reconocimiento, ya sea de fonemas, sílabas, palabras u oraciones completas para lograr transmitir el mensaje que se tiene la intención de decir sin que llegue realmente a ser expresado como una señal acústica. La figura 1 clasifica los enfoques de las SSI dependiendo de la fase de producción del habla de donde se esté tomando la señal a analizar. Vale la pena anotar que, básicamente, dependiendo del tipo de señal que se extraiga, se implementará una interfaz específica y será necesario realizar un acondicionamiento de señal adecuado para cada uno de los casos, a fin de conseguir el reconocimiento del habla silenciosa [68] [69].

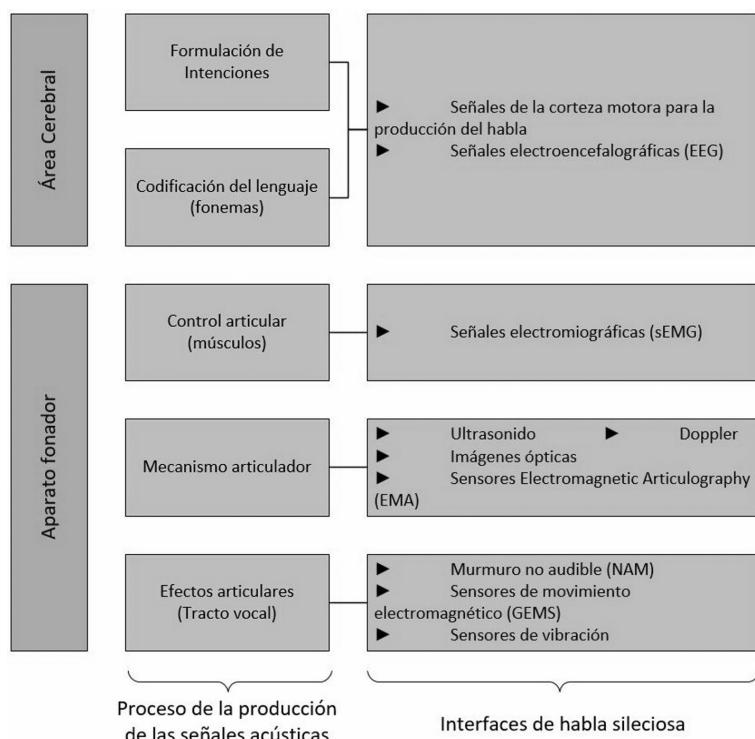


Figura 1. Enfoques de los SSIs durante la producción de la voz.

Son muchas las aplicaciones actuales de este tipo de interfaces, que van desde su utilidad en medicina de recuperación en pacientes que han sido sometidos a alguna intervención del aparato fonador, como por ejemplo una laringectomía y que requieren restablecer su habilidad comunicativa; pasando por los avances que se realizan en sistemas de comunicación humano-máquina, hasta dispositivos que permitan la interlocución en ambientes extremadamente ruidosos o en donde se requiere que el mensaje enviado presente altas condiciones de seguridad.

Este artículo pretende realizar una corta pero comprensible reseña de los métodos de adquisición y procesamiento del habla sub-vocal, los avances que se han realizado en dispositivos de comunicación silenciosa, a nivel mundial y especialmente en Colombia y Latinoamérica, mostrando las limitaciones actuales de los sistemas de habla silenciosa y las perspectivas del trabajo que se desarrollará en los próximos años en este campo. En [2], los autores realizan un detallado estudio de las técnicas actuales para el reconocimiento del habla sub-vocal, basándose en diferentes tipos de sensores, evaluando las ventajas, portabilidad y nivel de alcance comercial de cada método, haciendo una comparación entre estos, que se muestra resumida en la tabla 1. A nivel local, y especialmente, en referencia a los avances realizados para la captura y clasificación de fonemas y palabras del idioma español, los métodos que se han desarrollado principalmente son la adquisición de señales por electromiografía de superficie, como se referencia en [3] y [4], y en menor proporción por medio de encefalogramas [5] e imágenes ultrasónicas. En la siguiente sección se presenta una breve introducción a cada uno de estos métodos, para dar paso a las aplicaciones que tienen en la actualidad este tipo de interfaces.

## 2. **Métodos de adquisición y procesamiento del habla sub-vocal**

La tabla 1 describe los métodos más populares de SSI y hace una comparación entre ellos, destacando sus ventajas y posibilidad de comercialización a corto plazo, como se presenta en [2]. De acuerdo con la figura 1, la primera señal que es posible captar desde el momento en que se produce la intención de pronunciar una palabra es la proveniente de la corteza cerebral, donde se aloja el proceso del habla y establecen los comandos para la comunicación y la expresión oral. Para adquirir estas señales, típicamente se utilizan varios electrodos de alta resolución que se ubican en la parte izquierda del cerebro, donde se obtienen las señales EEG relacionadas con la producción del habla [5]. Mientras en algunos trabajos como en [6] se utiliza un casco que contiene un gel conductor y se fija a la cabeza del paciente para la apropiada ubicación de los electrodos, otras investigaciones como la presentada en [7] utilizan señales electrocorticográficas (ECOG), que se toman directamente sobre la superficie del cerebro, lo que evita la atenuación de las señales debido al hueso del cráneo y la posible interferencia de ruido causado por el cuero cabelludo. La principal desventaja que presenta dicha técnica es que es muy invasiva, por lo que su aplicación es frecuente en medicina en pacientes con condiciones médicas críticas. Las características para el reconocimiento del habla silenciosa son diferentes de las tomadas para el reconocimiento del habla acústica [6]. Los métodos de adquisición también son diferentes; mientras en el habla normal la toma de datos se realiza sobre un canal a 16kHz, las ondas cerebrales se adquieren a través de 16 canales a 300Hz, cada uno.

El objetivo de los BIC (Brain-Computer Interface, por sus siglas en inglés) es traducir los pensamientos o intenciones de un sujeto dado en una señal de control para operar dispositivos tales como computadores, sillas de ruedas o prótesis [8]. Frecuentemente, este tipo de interfaz requiere que el usuario explícitamente manipule su actividad cerebral, la cual se usa entonces para controlar la señal de un dispositivo. Generalmente se necesita un proceso de aprendizaje que puede tomar varios meses. Este es el caso de una prótesis de habla para individuos con impedimentos severos de comunicación [9], donde se propone un esquema de control de señales EEG, para ser implementada en pacientes que sufren esclerosis lateral amiotrofia avanzada (ELA).

**Tabla I.** Comparación de las tecnologías actuales

Tipo de interfaz	Trabaja de forma silenciosa	Funciona en entorno ruidoso	Rehabilitación	Tipo de procedimiento	Posibilidad de comercialización	Costo de implementación
Electrodos EEG	Sí	Sí	Muy utilizado en pacientes con condiciones médicas críticas (en especial ECoG)	Puede ser invasivo	No es invasivo	Medio
Electrodos EMG	Sí	Sí	En pacientes que han sido sometidos a una laringectomía	No es invasivo	Sí	Medio
Imágenes ultrasónicas	Sí	Sí	En pacientes que han sido sometidos a una laringectomía	No es invasivo	Sí	Medio/alto
Micrófono de murmullo no audible (NAM)	No completamente	Presenta interferencias por ruido proveniente principalmente del movimiento del paciente	No es usualmente utilizado para fines médicos. Sí para fines comerciales	No es invasivo	Sí	Bajo
Electromagnéticos o de vibración	No completamente	Sí	No es utilizado para fines médicos	No es invasivo	Sí	Medio/alto

Algunos de los avances realizados en el último año sugieren que es posible identificar lenguaje imaginario [10], inicialmente para la identificación de dos sílabas. Las aplicaciones EEG para el procesamiento de señales biológicas va más allá del reconocimiento de la intención de habla, trabajos como [11] se enfocan en identificar la intención humana a través de una interfaz humano-máquina. De igual forma, en [12] se propone un nuevo enfoque al problema de lenguaje silencioso por medio de una interfaz cerebro-computador (BCI), basada en micro-electrodos intracorticales para predecir la información que se destina al discurso, directamente desde la actividad de las neuronas que están relacionadas con la producción del habla.

Otro de los métodos que ha presentado un amplio desarrollo es la adquisición de señales por electromiografía de superficie (sEMG, por sus siglas en inglés, surface Electromyography). La utilización de esta tecnología consiste en registrar la actividad eléctrica del músculo, la cual es capturada por electrodos de superficie (es decir, no implantados), produciendo señales acordes a la vibración o movimiento del aparato fonador [13], [14]. Las señales eléctricas son producidas por la fibra muscular, que es activada por el sistema nervioso central, generando una pequeña descarga de corriente eléctrica debida a los flujos iónicos a través de las membranas de las células musculares. Después de adquirirse las señales, se realiza un proceso de amplificación para tener señales que puedan ser utilizadas en la reproducción de la voz [15], [16], [17]. Cuando se utiliza esta técnica en interfaces que pretenden instrumentar el reconocimiento del lenguaje en español, se busca utilizar el enfoque de reconocimiento de sílabas [18], ya que es un modo común y natural de dividir las palabras en español, lo que presenta una diferencia notoria con las interfaces desarrolladas para el reconocimiento del idioma Inglés, pues dicho idioma se divide de acuerdo a los golpes de sonido y no en sílabas. Este es el caso de la investigación realizada en [3], donde la interfaz de lenguaje silencioso desarrollada se basa en el reconocimiento de sílabas, en vez de fonemas o palabras, obteniendo una efectividad del 70 %, utilizando un árbol de decisiones en el reconocimiento de 30 monosílabos del español. Igualmente en [4], se realiza el análisis de las señales electromiográficas de los músculos mientras el paciente pronuncia fonemas correspondientes a 30 monosílabos diferentes. Para ello, entrenaron un clasificador a fin de reconocer las sílabas con una efectividad del 71 %. En [19], los autores conjugan las señales de habla normal, de susurro y de habla silenciosa, todas obtenidas

por sMEG, para su análisis y comparación, mostrando que las tasas de reconocimiento en el lenguaje silencioso mejoran en porcentaje considerable.

Cuando el proceso de producción de voz llega hasta el aparato fonador se presenta la posibilidad de implementar nuevas interfaces basadas en las vibraciones de este mismo. Tal es el caso de las adquisición de señales por micrófonos de Murmuro no Audible, NAM, señales de ultrasonido [20], [21] y ópticas [22], [23], [24] o utilizando sensores magnéticos [25], o sensores de vibración [26]. El reconocimiento del soplo no audible NAM (por las siglas en inglés de Non-audible Murmur), es una de las interfaces del lenguaje silencioso más prometedoras para comunicación hombre-máquina. El sonido de baja amplitud generado por el flujo de aire en la laringe puede ser detectado usando el micrófono de NAM [27], [28]. La principal desventaja de este dispositivo es la pérdida de las altas frecuencias de la señal debido al medio de transmisión de las mismas, lo cual causa que las señales detectadas no sean lo suficientemente claras [29]. Un micrófono NAM comprende un micrófono condensador (ECM, Electret Condenser Microphone, por sus siglas en inglés), cubierto de un polímero suave, como la silicona o el elastómero de uretano, que proporcionan una mejor impedancia y ayudan al contacto suave con el tejido del cuello. La sensibilidad del micrófono de acuerdo al material utilizado (silicona o elastómero de uretano) en 1kHz están entre – 41 y 58 dB. Como se observa en la figura 2, este se ubica en el cuello cerca de la oreja. En [30] se registra un estudio de la propagación del sonido desde el tracto vocal hasta la superficie del cuello con el objetivo de mejorar la claridad de las señales tipo NAM obtenidas.

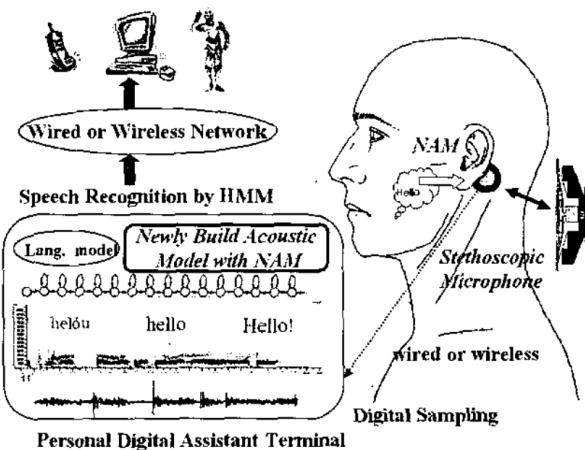


Figura 2. Normas relevantes en Colombia sobre calidad de potencia.

El principal método para el reconocimiento de señales es el entrenamiento de un modelo acústico basado en los modelos ocultos de Markov (HMM, Hidden Markov Model, por sus siglas en inglés), como se explica en [31]. Más específicamente, en el área del habla sub-vocal, HMM se ha utilizado para la conversión y comparación estadística de la articulación silenciosa del aparato fonador, con vocablos de una base de datos guardados con anterioridad [61]. Basado en este mismo modelo, se presenta en [32] un nuevo método de reconocimiento NAM, el cual requiere solamente una pequeña cantidad de datos para el entrenamiento de HMMs y está basado en adaptación supervisada e iteración adaptativa. Con el fin de mejorar la claridad de las señales NAM, en [33] los autores presentan los resultados de un escaneo por imágenes de resonancia magnética del tracto vocal para ser aplicadas al estudio del mecanismo

de producción del NAM y compararlo con el mecanismo de producción de habla normal. Uno de los trabajos más recientes hechos sobre el análisis y reconocimiento de las señales tipo NAM es el mostrado en [34]. Aquí los autores realizan el reconocimiento de los fonemas japoneses e incluyen el uso de señales provenientes del habla normal con las señales tipo NAM para un mejor procesamiento. Lo anterior debido a que las señales NAM muestreadas resultan débiles, por lo que requieren un proceso de amplificación antes de ser analizadas por herramientas de reconocimiento del habla. Este es un nuevo enfoque para métodos de reconocimientos de señales NAM que incluye la implementación de un modelo acústico y otro de lenguaje, así como la utilización de datos de habla normal transformados en datos NAM, como es posible observar en [35]. Para combatir el ruido causado, entre otras cosas por el movimiento del hablante, en [36] se plantea el uso de un detector de señal estéreo junto a dos micrófonos NAM para la supresión del ruido por medio de la separación ciega de fuentes y de la sustracción espectral que se realiza en cada uno de los canales.

Otro método empleado para el procesamiento de las señales de habla sub-vocal se presenta en [62], donde los autores filtran con la Transformada Wavelet Discreta, las señales tomadas con electrodos superficiales colocados en la garganta, para que a continuación se extraigan las principales características de las señales y se clasifiquen con redes neuronales. A fin de analizar las señales de habla sub-vocal, también se han usado algoritmos como las series de Fourier-Bassel, que separan los formantes de la señal de voz y luego, usando la transformada de Hilbert, se calcula la frecuencia instantánea de cada formante separado [63]. Asimismo, utilizando la transformada fraccionada del coseno, es posible analizar la información de habla, reconociendo las palabras ocultas en la señal de habla enviada en un sistema transmisor-receptor [64].

### **3. Avances en las tecnologías actuales**

Los anteriores enfoques para la adquisición del lenguaje silencioso han permitido el desarrollo de tecnologías tales como micrófonos, interfaces de reconocimiento del habla, implantes cocleares, entre otros; que abarcan un gran número de aplicaciones en diferentes campos, como se ha dicho: medicina, comunicaciones, seguridad, etc. Muchos de estos avances se han desarrollado con el fin de subsanar las desventajas que presentan las interfaces de lenguaje silencioso, como la poca portabilidad de los sistemas desarrollados, su viabilidad en aplicaciones en entornos reales al ser un sistema basado netamente en la información proveniente de los sensores empleados, sin poseer realimentación de audio [37], así como la degradación e interferencia de las señales en condiciones de ruido extremas. Los sensores no acústicos [38, 39], tales como los radares de microondas, los de vibración de la piel y los sensores de conducción ósea, proporcionan la posibilidad de medir la excitación de la glotis y del tracto vocal siendo totalmente inmunes a las interferencias acústicas.

Otra de las mejoras implementadas en la adquisición del habla sub-vocal es la inclusión de implantes cocleares para mejorar la discriminación individual de palabras y las habilidades del entendimiento del habla silenciosa, como se describe en [40], [41] y [42]. En [43] se combina un procesador coclear con un circuito integrado al tracto vocal para crear una “realimentación del habla”, que puede aplicarse a prótesis de escucha o para mejorar el reconocimiento del habla en ambientes extremadamente ruidosos. También se investiga su implementación en una interfaz hombre máquina o en sintetizadores del habla usando un dispositivo que simule el

tracto vocal humano. El sistema de realimentación de voz es explicado por Keng y Hoong con detalle en [44] y en [45]. Igualmente, ese último trabajo aplica el concepto de realimentación de las señales de habla para la disminución del ruido de la señal y, así, se sintetice mejor y de forma clara.

## 4. Aplicaciones actuales del habla sub-vocal

El lenguaje silencioso o habla sub-vocal ha sido investigado durante décadas y ahora es una tecnología lo suficientemente madura para ser usada en una amplia variedad de aplicaciones, como en el área de las comunicaciones, en sistemas de información telefónica, en programas de traducción entre idiomas y de dictado, que pueden ser implementados en un nivel superior, más rápido y con mejor calidad en el reconocimiento de información que con las interfaces de lenguaje acústico [46]. En el área de la medicina, especialmente en rehabilitación del tracto vocal, como se muestra en [47] y [48], cuando un paciente ha sido sometido a una laringectomía debido a un accidente o al cáncer de laringe, no puede producir el sonido de la voz de una manera convencional debido a que sus cuerdas vocales han sido retiradas, por lo que se requiere un método alternativo de habla para producir la voz usando fuentes que generen el sonido de manera especial, sin necesidad de hacer vibrar sus cuerdas vocales. La voz producida por medio de este método se conoce como voz esofágica (que no es producida en la laringe). La utilización del habla sub-vocal en pacientes que han sido sometidos a una laringectomía total se presenta en [49]. El reconocimiento de la voz no laríngea y el incremento de la calidad de las señales producidas por una laringe electrónica usando métodos de reconocimiento de patrones son temas estudiados en [50]. Otro tipo de pacientes son aquellos que sufren de parálisis severas que les impiden realizar el proceso completo de producción de voz, pero sus señales en la corteza cerebral permiten la aplicación de interfaces BCI para hacer posible su comunicación con el entorno, como se muestra en [51] y [52]. En [12] se utilizan Electrodo Neurotrópicos para el desarrollo de la BCI, los cuales proveen grabaciones de las señales neurológicas útiles durante aproximadamente cuatro años, una propiedad necesaria para pacientes con parálisis que deben convivir mucho tiempo con este tipo de interfaces. Dentro de estas mismas interfaces se han elaborado sintetizadores de voz, que convierten las señales del habla sub-vocal en fonemas pronunciados por un dispositivo electrónico [65]. La utilidad que se le ha dado a estos desarrollos ha permitido profundizar más en los métodos que se han utilizado para captar las señales, el procesamiento digital y analógico de estas señales y los dispositivos utilizados para convertirlas en señales acústicas.

## 5. Trabajos realizados

En esta sección se citarán de forma más detallada, algunos de los trabajos que han sido desarrollados utilizando el habla sub-vocal. Una de las investigaciones realizadas en el campo del habla sub-vocal fue elaborada por la NASA, quienes investigaron en la aplicación del sEMG en ambientes hostiles [53]. Los resultados obtenidos con ese trabajo demuestran un 74 % de precisión para la clasificación de quince palabras del lenguaje inglés, en un sistema de tiempo real donde fue aplicado a sujetos que están expuestos a 95dB nivel de ruido. En [54] se presenta un dispositivo con la capacidad de medir la distribución de la vibración de las cuerdas vocales en tiempo real. Pacientes con enfermedades de la laringe fueron tratados con este dispositivo y los resultados obtenidos describen que fue posible estimar la frecuencia de los armónicos de los sonidos. En [55] y [56] se presentan analizadores de voz que usan

electrodos EMG, micrófonos Electrec y acelerómetros, para tratar a pacientes con desórdenes de lenguaje y la implementación del habla sub-vocal para la fabricación de sintetizadores del lenguaje.

En cuanto a comunicaciones respecta, en [57] se muestra una SSI compuesta por un sistema de adquisición multimodal de información del habla llamado Ultraspeech [37]. El sistema graba imágenes ultrasónicas y ópticas sincrónicamente con la señal de audio, que pueden ser transmitidas en tiempo real hasta la ubicación del receptor del mensaje. Las aplicaciones de las SSIs también abarcan las teleconferencias [58], las terminales telefónicas y los teléfonos móviles que tenga disponible la adaptación de tales herramientas de lenguaje silencioso [59]. Algunas investigaciones muestran estudios detallados en los que se expone que las comunicaciones telefónicas pueden ser reconocidas con precisión mediante señales sEMG, cuyas características pueden ser reconocidas, presentando los primeros resultados en clasificaciones de fonemas y características fonéticas en telefonía [46], aplicando técnicas estáticas de conversión de voz [60]. Otro de los trabajos derivados del habla sub-vocal es la laringe electrónica; este dispositivo transforma las vibraciones articulares del aparato fonador, en sonidos de voz, utilizando un sintetizador de voz que apoya la restitución del habla en pacientes que han sufrido cáncer de laringe [65]. A nivel local, uno de los trabajos con habla sub-vocal se presenta en [62], donde se utilizaron sensores superficiales y herramientas matemáticas para captar las señales bioeléctricas y clasificarlas en seis palabras de del idioma español, dando como resultado un 70 % de acierto. Asimismo, grupos de investigación de las universidades de Colombia han presentado sus trabajos para la adquisición de habla subvocal a través de USB [66], con una ancho de banda de 20-500 Hz y reconocimiento de fonemas monosílabicos con redes neuronales, con un porcentaje superior del 80 % [67].

## 6. Conclusiones

El objetivo principal de este documento fue presentar una revisión, corta pero profunda, de los trabajos realizados sobre el desarrollo de las interfaces de habla silenciosa, evidenciando las ventajas que ofrece para mejorar la comunicación en ambientes hostiles o con un nivel alto de ruido y los beneficios que se generan en personas que sufren de problemas en el aparato fonador y que no pueden producir los sonidos del habla. Es observable que en los últimos diez años los procedimientos donde se aplican las SSIs utilizan sensores colocados sobre el cuerpo, que proporcionan detalladamente las señales bioeléctricas transmitidas desde el cerebro hasta el tracto vocal, las cuales pueden tomarse en cada una de las etapas de producción de habla. Los trabajos presentados demuestran que, dependiendo de la técnica que se utilice, es posible identificar las señales de los fonemas de un idioma específico, para luego comparar la información obtenida con modelos estadísticos y sistemas con inteligencia artificial basados en métodos utilizados para el reconocimiento de habla.

Siendo el habla sub-vocal una técnica novedosa para la comunicación entre personas y la interacción con aquellos que poseen trastornos de habla o pérdida del aparato fonador, se hace necesario que las nuevas tecnologías que contemplen la aplicación de las interfaces de habla silenciosa, para el reconocimiento de fonemas de la lengua española a nivel nacional, se basen en los avances que se han generado en el ámbito mundial, que fomente la investigación y proponga nuevas alternativas de comunicación entre hablantes del idioma español. Adicionalmente, las nuevas tecnologías que se generen pueden fabricarse con un costo menor y adecuarse a los requerimientos de la población particular.

## Referencias

- [1] J. Freitas, A. Teixeira, M. Sales y C. Bastos, "Towards a Multimodal Silent Speech Interface for European Portuguese", en *Speech Technologies*, I. Ipsic, Ed., ed. Rijeka, Croatia: InTech, 2011, pp. 125-150.
- [2] B. Denby, T. Schultz, K. Honda, T. Hueber, J. Gilbert, J. Brumberg, "Silent speech interfaces", *Speech Communication*, vol. 52, pp. 270-287, 2010.
- [3] E. Lopez-Larraz, O. Martinez, J. Antelis y J. Minquez, "Syllable-Based Speech Recognition Using EMG", en *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, United States, 2010, pp. 4699-4702.
- [4] E. Lopez-Larraz, O. Martinez, J. Antelis, J. Damborenea y J. Minguez, "Diseño de un sistema de reconocimiento del habla mediante electromiografía", *XXVII Congreso Anual de la Sociedad Española de Ingeniería Biomédica*, Cádiz, España, 2009, pp. 601-604.
- [5] A. A. Torres García, C. A. Reyes García y L. Villaseñor Pineda, "Hacia la clasificación de habla no pronunciada mediante electroencefalogramas (EEG)", *XXXIV Congreso Nacional de Ingeniería Biomédica*, Ixtapa-Zihuatanajo, Guerrero, Mexico, 2011.
- [6] M. Wester, "Ünspoken Speech", Diplomarbeit, Inst. Theoretical computer science, Universität Karlsruhe (TH), Karlsruhe, Germany, 2006.
- [7] P. Xiaomei, J. Hill y G. Schalk, "Silent Communication: Toward Using Brain Signals", *Pulse*, IEEE, vol. 3, pp. 43-46, 2012.
- [8] A. Porbadnik, M. Wester, J. P. Calliess y T. Schultz, ".EEG-based Speech Recognition - Impact of Temporal Effects", *BIOSIGNALS*, 2009, pp. 376-381.
- [9] C. S. DaSalla, H. Kambara, M. Sato y Y. Koike, "Single-trial classification of vowel speech imagery using common spatial patterns", *Neural Networks*, vol. 22, pp. 1334-1339, 2009.
- [10] K. Brigham y B. V. K. Vijaya Kumar, "Imagined Speech Classification with EEG Signals for Silent Communication: A Preliminary Investigation into Synthetic Telepathy", *Bioinformatics and Biomedical Engineering (iCBBE)*, 2010 4th International Conference on, 2010, pp. 1-4.
- [11] D. Suh-Yeon y L. Soo-Young, "Understanding human implicit intention based on frontal electroencephalography (EEG)", *Neural Networks (IJCNN)*, The 2012 International Joint Conference on, 2012, pp. 1-5.
- [12] J. S. Brumberg, A. Nieto-Catanon, P. R. Kennedy y F. H. Guenther, "Brain-computer interfaces for speech communication", *Speech Communication*, vol. 52, pp. 367-379, 2010.
- [13] T. Schultz y M. Wand, "Modeling coarticulation in EMG-based continuous speech recognition", *Speech Communication*, vol. 52, pp. 341-353, 2010.
- [14] J. A. G. Mendes, R. R. Robson, S. Labidi y A. K. Barros, "Subvocal Speech Recognition Based on EMG Signal Using Independent Component Analysis and Neural Network MLP", *Image and Signal Processing*, 2008. CISP '08. Congress on, 2008, pp. 221-224.
- [15] L. Maier-Hein, "Speech Recognition Using Surface Electromyography", Diplomarbeit, Universitat Karlsruhe, Karlsruhe, 2005.
- [16] G. S. Meltzner, G. Colby, Y. Deng y J. T. Heaton, "Signal acquisition and processing techniques for sEMG based silent speech recognition", *Engineering in Medicine and Biology Society, EMBC*, 2011 Annual International Conference of the IEEE, 2011, pp. 4848-4851.
- [17] J. Freitas, A. Teixeira y M. Sales Dias, "Towards a silent speech interface for portuguese surface electromyography and the nasalinity challenge", *International Conference on Bio-inspired Systems and Signal Processing, BIOSIGNALS 2012*, February 1, 2012 - February 4, 2012, Vilamoura, Algarve, Portugal, 2012, pp. 91-100.
- [18] T. Schultz, "ICCHP keynote: Recognizing silent and weak speech based on electromyography", 12th International Conference on Computers Helping People with Special Needs, ICCHP 2010, July 14, 2010 - July 16, 2010, Vienna, Austria, 2010, pp. 595-604.
- [19] M. Janke, M. Wand y T. Schultz, ".A Spectral Mapping Method for EMG-based Recognition of Silent Speech", *B-Interface 2010 - Proceedings of the 1st International Workshop on Bio-inspired Human-Machine Interfaces and Healthcare Applications*, Valencia, España, 2010, pp. 22-31.
- [20] B. Denby, Y. Oussar, G. Dreyfus y M. Stone, "Prospects for a Silent Speech Interface using Ultrasound Imaging", *Acoustics, Speech and Signal Processing*, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on, 2006, pp. I-I.

- [21] B. Denby y M. Stone, "Speech synthesis from real time ultrasound images of the tongue", *Acoustics, Speech, and Signal Processing*, 2004. Proceedings. (ICASSP '04). IEEE International Conference on, 2004, pp. I-685-8 vol.1.
- [22] T. Hueber, G. Chollet, B. Denby, G. Dreyfus y M. Stone, "Continuous-Speech Phone Recognition from Ultrasound and Optical Images of the Tongue and Lips", *International Speech Communication Association INTERSPEECH 2007*, 8th Annual Conference of the, Antwerp, Belgium, 2007, pp. 658-661.
- [23] T. Hueber, G. Chollet, B. Denby, G. Dreyfus y M. Stone, "Phone Recognition from Ultrasound and Optical Video Sequences for a Silent Speech Interface", *International Speech Communication Association INTERSPEECH 2008*, 9th Annual Conference of the, Brisbane. Australia, 2008, pp. 2032-2035.
- [24] T. Hueber, E. L. Benaroya, G. Chollet, B. Denby, G. Dreyfus y M. Stone, "Development of a silent speech interface driven by ultrasound and optical images of the tongue and lips", *Speech Communication*, vol. 52, pp. 288-300, 2010.
- [25] J. M. Gilbert, S. I. Rybchenko, R. Hofe, S. R. Ell, M. J. Fagan, R. K. Moore y P. Green, "Isolated word recognition of silent speech using magnetic implants and sensors", *Medical Engineering Physics*, vol. 32, pp. 1189-1197, 2010.
- [26] R. Hofe, S. R. Ell, M. J. Fagan, J. M. Gilbert, P. D. Green, R. K. Moore y S. I. Rybchenko, "Small-vocabulary speech recognition using a silent speech interface based on magnetic sensing", *Speech Communication*, vol. 55, pp. 22-32, 2012.
- [27] T. Toda, "Statistical approaches to enhancement of body-conducted speech detected with non-audible murmur microphone", *Complex Medical Engineering (CME)*, 2012 ICME International Conference on, 2012, pp. 623-628.
- [28] D. Babani, T. Toda, H. Saruwatari y K. Shikano, "Acoustic model training for non-audible murmur recognition using transformed normal speech data", *Acoustics, Speech and Signal Processing (ICASSP)*, 2011 IEEE International Conference on, 2011, pp. 5224-5227.
- [29] Y. Nakajima, H. Kashioka, K. Shikano y N. Campbell, "Non-audible murmur recognition input interface using stethoscopic microphone attached to the skin", *Acoustics, Speech, and Signal Processing*, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on, 2003, pp. V-708-11 vol.5.
- [30] Y. Nakajima, H. Kashioka, N. Campbell y K. Shikano, "Non-Audible Murmur (NAM) Recognition", *IEICE - Trans. Inf. Syst.*, vol. E89-D, pp. 1-4, 2006.
- [31] V.-A. Tran, G. Bailly, H. Leevenbruck y T. Toda, "Improvement to a NAM-captured whisper-to-speech system", *Speech Communication*, vol. 52, pp. 314-326, 2010.
- [32] S. Shimizu, M. Otani, T. Hirahara, "Frequency characteristics of several non-audible murmur (NAM) microphones", *Acoustical Science and Technology*, vol. 30, pp. 139-142, 2009.
- [33] T. Hirahara, M. Otani, S. Shimizu, T. Toda, K. Nakamura, Y. Nakajima y K. Shikano, "Silent-speech enhancement using body-conducted vocal-tract resonance signals", *Speech Communication*, vol. 52, pp. 301-313, 2010.
- [34] M. Otani, T. Hirahara, S. Shimizu y S. Adachi, "Numerical simulation of transfer and attenuation characteristics of soft-tissue conducted sound originating from vocal tract", *Applied Acoustics*, vol. 70, pp. 469-472, 2009.
- [35] P. Heracleous, V.-A. Tran, T. Nagai y K. Shikano, "Analysis and Recognition of NAM Speech Using HMM Distances and Visual Information", *Audio, Speech, and Language Processing*, IEEE Transactions on, vol. 18, pp. 1528-1538, 2010.
- [36] S. Ishii, T. Toda, H. Saruwatari, S. Sakti y S. Nakamura, "Blind noise suppression for Non-Audible Murmur recognition with stereo signal processing", *Automatic Speech Recognition and Understanding (ASRU)*, 2011 IEEE Workshop on, 2011, pp. 494-499.
- [37] V. Florescu, L. Crevier-Buchman, B. Denby, T. Hueber, A. Colazo-Simon, C. Pillot-Loiseau, P. Roussel, C. Gendrot y S. Quattrociocchi, "Silent vs Vocalized Articulation for a Portable Ultrasound-Based Silent Speech Interface", *IEEE 18th Signal Processing and Communications Applications Conference*, Diyarbakir, Turkey, 2010.
- [38] S. Li, J. Q. Wang, J. XiJing y T. Liu, "Nonacoustic Sensor Speech Enhancement Based on Wavelet Packet Entropy", *Computer Science and Information Engineering*, 2009 WRI World Congress on, 2009, pp. 447-450.
- [39] T. F. Quatieri, K. Brady, D. Messing, J. P. Campbell, M. S. Brandstein, C. J. Weinstein, J. D. Tardelli y P. D. Gatewood, "Exploiting nonacoustic sensors for speech encoding", *Audio, Speech, and Language Processing*, IEEE Transactions on, vol. 14, pp. 533-544, 2006.

- [40] L. Xiaoxia, N. Kaibao, F. Karp, K. L. Tremblay y J. T. Rubinstein, "Characteristics of stimulus artifacts in EEG recordings induced by electrical stimulation of cochlear implants", Biomedical Engineering and Informatics (BMEI), 2010 3rd International Conference on, 2010, pp. 799-803.
- [41] D. Smith y D. Burnham, "Facilitation of Mandarin tone perception by visual speech in clear and degraded audio: Implications for cochlear implants", The Journal of the Acoustical Society of America, vol. 131, pp. 1480-1489, 2012.
- [42] F. Lin, W. W. Chien, D. M. Clarrett, J. K. Niparko y H. W. Francis, "Cochlear Implantation in Older Adults", Medicine, vol. 91, pp. 229-241, 2012.
- [43] W. Keng Hoong, L. Turicchia y R. Sarpeshkar, "An Articulatory Silicon Vocal Tract for Speech and Hearing Prostheses", Biomedical Circuits and Systems, IEEE Transactions on, vol. 5, pp. 339-346, 2011.
- [44] W. Keng Hoong, L. Turicchia y R. Sarpeshkar, "An Articulatory Speech-Prosthesis System", Body Sensor Networks (BSN), 2010 International Conference on, 2010, pp. 133-138.
- [45] W. Keng Hoong, L. Turicchia y R. Sarpeshkar, "A speech locked loop for cochlear implants and speech prostheses", Applied Sciences in Biomedical and Communication Technologies (ISABEL), 2010 3rd International Symposium on, 2010, pp. 1-2.
- [46] M. Wand y T. Schultz, "Analysis of phone confusion in EMG-based speech recognition", Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, 2011, pp. 757-760.
- [47] M. J. Fagan, S. R. Ell, J. M. Gilbert, E. Sarrazin y P. M. Chapman, "Development of a (silent) speech recognition system for patients following laryngectomy", Medical Engineering & Physics, vol. 30, pp. 419-425, 2008.
- [48] N. Alves, T. H. Falk y T. Chau, "A novel integrated mechanomyogram-vocalization access solution", Medical Engineering and Physics, vol. 32, pp. 940-944, 2010.
- [49] H. Doi, K. Nakamura, T. Toda, H. Saruwatari y K. Shikano, "An evaluation of alaryngeal speech enhancement methods based on voice conversion techniques", Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, 2011, pp. 5136-5139.
- [50] J. S. Brumberg, P. R. Kennedy y F. H. Guenther, "Artificial speech synthesizer control by brain-computer interface", International Speech Communication Association INTERSPEECH 2009, 10th Annual Conference of the, Brighton, Reino Unido, 2009.
- [51] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson y T. M. Vaughan, "Brain-computer interface technology: a review of the first international meeting", Rehabilitation Engineering, IEEE Transactions on, vol. 8, pp. 164-173, 2000.
- [52] W. Jun, A. Samal, J. R. Green y F. Rudzicz, "Sentence recognition from articulatory movements for silent speech interfaces", Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, 2012, pp. 4985-4988.
- [53] M. Braukus y J. Bluck. (2004, March 17). NASA Develops System To Computerize Silent, "Subvocal Speech". Disponible en: [http://www.nasa.gov/home/hqnews/2004/mar/HQ\\_04093\\_subvocal\\_speech.html](http://www.nasa.gov/home/hqnews/2004/mar/HQ_04093_subvocal_speech.html)
- [54] I. Ishii, S. Takemoto, T. Takaki, M. Takamoto, K. Imon y K. Hirakawa, "Real-time laryngoscopic measurements of vocal-fold vibrations", Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, 2011, pp. 6623-6626.
- [55] A. Carullo, A. Penna, A. Vallan, A. Astolfi y P. Bottalico, "A portable analyzer for vocal signal monitoring", Instrumentation and Measurement Technology Conference (I2MTC), 2012 IEEE International, 2012, pp. 2206-2211.
- [56] M. Janke, M. Wand, K. Nakamura y T. Schultz, "Further investigations on EMG-to-speech conversion", Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, 2012, pp. 365-368.
- [57] T. Hueber, G. Chollet, B. Denby y M. Stone, "Acquisition of Ultrasound, Video and Acoustic Speech Data for a Silent-Speech Interface Application", 8th International Seminar on Speech Production, Strasbourg, Francia, 2008.
- [58] H. Gamper, S. Tervo y T. Lokki, "Speaker Tracking for Teleconferencing via Binaural Headset Microphones", Acoustic Signal Enhancement; Proceedings of IWAENC 2012; International Workshop on, pp. 1-4, 2012.
- [59] K. A. Yuksel, S. Buyukbas y S. H. Adali, "Designing mobile phones using silent speech input and auditory feedback", 13th International Conference on Human-Computer Interaction with Mobile Devices and Services, Mobile HCI 2011, August 30, 2011 - September 2, 2011, Stockholm, Sweden, 2011, pp. 711-713.

- [60] T. Toda, M. Nakagiri y K. Shikano, "Statistical Voice Conversion Techniques for Body-Conducted Unvoiced Speech Enhancement", *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, pp. 2505-2517, 2012.
- [61] T. Hueber y G. Bailly, "Statistical conversion of silent articulation into audible speech using full-covariance HMM", *Computer Speech and Language*, pp. 1-20, Article in Press, 2015.
- [62] L. Mendoza, J. Peña Martínez, L. A. Muñoz-Bedoya y H. J. Velandia-Villamizar, "Procesamiento de Señales Provenientes del Habla Subvocal usando Wavelet Packet y Redes Neuronales", *TecnoLógicas*, vol. Edición Especial, pp.655-667, 2013.
- [63] P. S. Rathore, A. Boyat y B. K. Joshi, "Speech signal analysis using Fourier-Bessel Expansion and Hilbert Transform Separation Algorithm", *Signal Processing, Computing and Control (ISPCC), 2013 IEEE International Conference on*, Septiembre 26-28, pp. 1-4, 2013.
- [64] Y. Tao y Y. Bao, "Speech Information Hiding using Fractional Cosine Transform", en 1st International Conference on Information Science and Engineering, Diciembre 26-28, pp. 1864-1867, 2009.
- [65] A. Razo, "Sistema para mejorar la calidad de la voz esofágica", disertación para M.Sc., Sección de estudios de posgrado e investigación, Instituto Politécnico Nacional, México D.F.
- [66] Grupo de Ingeniería Biomédica de la Universidad de Pamplona (GIBUP), (2010, mayo 3) "Adquisición por USB del habla subvocal". Disponible en: <https://gibup.wordpress.com/2010/05/03/adquisicion-por-usb-del-habla-subvocal/>
- [67] G. Buitrago, "Reconocimiento de patrones del habla subvocal utilizando algoritmos evolutivos", Informe de Investigación, Grupo de Investigación GAV, Programa de Ingeniería en Mecatrónica, Universidad Militar Nueva Granada, Bogotá D.C., Colombia, 2013.
- [68] V. Havas, A. Gabarrós, M. Juncadella, X. Rifa-Ríos, G. Plans, J. J. Acebes, R. de Diego Balaquer y A. Rodriguez-Fornells, "Electrical stimulation mapping of nouns and verbs in Broca's area", *Brain and Language*, vol. 145-146, pp. 53-63, 2015.
- [69] J. Gutiérrez, E. N. Gama, D. Amaya y O. Avilés, "Desarrollo de interfaces para la detección del habla subvocal", *Tecnura*, vol. 17, No. 37, pp. 138-152, 2013.

---

### **Erika Nathalia Gamma**

Nació en Bogotá D.C., Colombia. Es Ingeniera en Mecatrónica de la Universidad Militar Nueva Granada, de Bogotá, Colombia. Se desempeñó como Ingeniera en Colvista SAS durante 1 años. Posteriormente, ejerció como ingeniería en Cosmos Sistemas Integrados durante 1 año. Luego se desempeñó como Joven Investigadora en la Universidad Militar Nueva Granada desde el 2011 para el apoyo en investigación del grupo GAV.  
e-mail: gav@unimilitar.edu.co

---

### **Darío Amaya Hurtado**

Nació en Quindío, Colombia. Es Ingeniero Electrónico de la Universidad Antonio Nariño, de Bogotá D.C., Colombia. Obtuvo su título de Maestría en teleinformática en la Universidad Distrital Francisco José de Caldas de Bogotá D.C., Colombia. Obtuvo su PhD en la Universidad Estadual de Campinas, de São Paulo, Brasil. Se desempeñó como docente en fundación universidad central durante 5 años. Actualmente se desempeña como profesor en el área de Automatización Industrial y sensores en la Universidad Militar Nueva Granada de Bogotá D.C., Colombia, y pertenece como investigador al grupo GAV donde realiza estudios sobre energías alternativas y habla Sub-Vocal.  
e-mail: dario.amaya@unimilitar.edu.co

---

### **Olga Lucia Ramos Sandoval**

Nació en Bogotá D.C., Colombia. Es Ingeniera Electrónica de la Universidad Antonio Nariño, de Bogotá, Colombia. Obtuvo su título de Maestría en teleinformática en la Universidad Distrital Francisco José de Caldas de Bogotá D.C., Colombia. Se encuentra realizando su PhD en la Universidad Distrital Francisco José de Caldas, de Bogotá D.C., Colombia. Se desempeñó como docente de tiempo completo en la Universidad Manuela Beltrán durante 4 años. Actualmente se desempeña como profesor en el área de Automatización Industrial en la Universidad Militar Nueva Granada de Ciudad, País, y pertenece como investigador al grupo GAV donde realiza estudios sobre habla Sub-Vocal.  
e-mail: olga.ramos@unimilitar.edu.co