

ANÁLISIS DE LA INFLUENCIA DEL RUIDO PROVOCADO POR SISTEMAS DE
ADQUISICIÓN DE IMÁGENES DE RESONANCIA MAGNÉTICA (MRI) SOBRE
LOS PARÁMETROS ACÚSTICOS DE LA VOZ

LUIS CARLOS ESPELETA CADENA

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICO-MECÁNICAS
ESCUELA DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA Y DE
TELECOMUNICACIONES
BUCARAMANGA
2019

ANÁLISIS DE LA INFLUENCIA DEL RUIDO PROVOCADO POR SISTEMAS DE
ADQUISICIÓN DE IMÁGENES DE RESONANCIA MAGNÉTICA (MRI) SOBRE
LOS PARÁMETROS ACÚSTICOS DE LA VOZ

LUIS CARLOS ESPELETA CADENA

TRABAJO DE GRADO PARA OPTAR EL TÍTULO DE INGENIERO
ELECTRÓNICO

DIRECTOR

FRANKLIN ALEXANDER SEPÚLVEDA SEPÚLVEDA
PhD EN INGENIERÍA

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISCO-MECÁNICAS
ESCUELA DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA Y DE
TELECOMUNICACIONES
BUCARAMANGA

2019

DEDICATORIA,

A la memoria de:

Mi padre Carlos Enrique Espeleta Garrido (20/05/57 - 09/01/12)

Y mis tías:

Marina del Carmen Espeleta Garrido (16/07/51 - 18/12/15)

Celestina Espeleta Garrido (06/01/48 - 14/09/16)

TABLA DE CONTENIDO

INTRODUCCIÓN	16
1. OBJETIVOS Y ALCANCE:	20
1.1 OBJETIVOS GENERALES	20
1.2 OBJETIVOS ESPECÍFICOS	20
1.3 ALCANCE	20
2. MARCO TEÓRICO.....	21
2.1 PRINCIPIOS BÁSICOS DE LOS SISTEMAS DE MRI	21
2.2 COMPONENTES DE UN SISTEMA DE MRI	23
2.3 CARACTERÍSTICAS DEL RUIDO ACÚSTICO DE MRI	27
3. MÉTODO.....	31
3.1. BASE DE DATOS USC-TIMIT	32
3.1.1 Adquisición de la señal de audio.	34
3.1.2 Reducción del ruido acústico en la base de datos USCM-TIMIT.	36
3.2 REPRESENTACIÓN DE LA SEÑAL DE VOZ	37
3.2.1 Formantes.	37
3.2.2 Frecuencia Fundamental (F_0).	38
3.2.3 Jitter (relativo).....	39
3.2.4 Shimmer.	40
3.2.5 Mean Harmonic to Noise Ratio (HNR).....	40
3.2.6 Envolverte espectral por LPC (Linear Prediction Coefficients).	41
3.2.7 Mel Frequency Cepstrales Coefficients (MFCC) y envolverte espectral.	43
3.2.8 Long Term Average Spectrum (LTAS).	45
3.3 MEDIDAS DE COMPARACIÓN DE SEÑALES DE VOZ	47
3.3.1 Medidas log espectrales.....	48
3.3.2 Coeficiente de correlación espectral de Pearson (P).....	48
3.3.3 Log Likelihood Ratio (LLR).	49
3.4 MEDIDAS OBJETIVAS DE LA CALIDAD DE LA VOZ:	49
3.4.1 Medidas compuestas CSIG, CBAK, COVL.	49
3.4.2 Perceptual Evaluation of Speech Quality (PESQ).	51

4. RESULTADOS Y DISCUSIÓN	53
CONCLUSIONES:.....	68
BIBLIOGRAFÍA:	71
ANEXOS	77

LISTA DE TABLAS

Tabla 1. Participantes del estudio-Detalles demográficos.....	33
Tabla 2. Descripción de la escala de las tres dimensiones (SIG, BAK, OVL) usada en la metodología ITU-T P8.35.	50
Tabla 3. Conjunto de coeficientes a_0 , a_1 , a_2 de PESQ para cada dimensión de calidad.....	52
Tabla 4. Resultados de las medidas para la evaluación objetiva de la calidad de la voz.....	53
Tabla 5. Resultado de las dos medidas log espectrales (ec 9) para comparar las tres representaciones espectrales.....	58
Tabla 6. Coeficientes de correlación de Pearson (ec 10) de las tres representaciones de la envolvente espectral.....	59

LISTA DE FIGURAS

Figura 1. Núcleos de Hidrogeno en ausencia y presencia de un campo magnético fuerte.	22
Figura 2. (a) Pulso RF aplicado. (b) Cuando se detiene el pulso RF.	23
Figura 3. : Escáner típico de MRI y sus componentes.	24
Figura 4. Las tres bobinas de gradiente.	25
Figura 5. Bobina de Radio Frecuencia (RF).	26
Figura 6. Fuentes de ruido y ubicación de los micrófonos en una sala de escaneo de MRI.	28
Figura 7. Método seguido durante la realización del proyecto de grado	31
Figura 8. Tractos vocales de los diez hablantes.	33
Figura 9. Diagrama de adquisición de audio de la base de datos USC-TIMIT.	35
Figura 10. Principio de operación de los MFO.	36
Figura 11. Método de cancelación del ruido.	37
Figura 12. Parámetros usados para el cálculo de los formantes con el software PRAAT.	38
Figura 13. Parámetros usados para el cálculo de Frecuencia fundamental con el software PRAAT.	39
Figura 14. Envolventes espectrales por LPC de 4 diferentes frases de la base de datos USCM-TIMIT.	42
Figura 15. Diagrama de bloques para la extracción de los MFCC.	43
Figura 16. Envolventes espectrales por MFCC de 4 diferentes frases de la base de datos USCM-TIMIT.	44
Figura 17. LTAS de 4 diferentes frases de la base de datos USCM-TIMIT.	46
Figura 18. Relación entre los resultados de csig y la escala SIG.	54
Figura 19. Relación entre los resultados de csig y la escala SIG.	54
Figura 20. Relación entre los resultados de covl y la escala OVL.	55
Figura 21. Resultados de la medida LLR.	55
Figura 22. Resultados de la medida PESQ.	56

Figura 23. Comparación entre los valores de PESQ y LLR	57
Figura 24. Comparación entre la medida rms log espectral (ec.9,p=2) de las tres representaciones de la envolvente espectral de cada hablante, con el umbral de distorsión.....	58
Figura 25. Boxplots de las medidas acústicas Fo, jitter, shimmer y HNR.....	60
Figura 26. Boxplot de los cuatros formantes (F1, F2, F3, F4).	62
Figura 27. Valor promedio NRMSE de cada medida acústica.....	67

LISTA DE ANEXOS

ANEXOS A. Frases de la base de datos USCM-TIMIT	77
---	----

RESUMEN

TITULO: ANÁLISIS DE LA INFLUENCIA DEL RUIDO PROVOCADO POR SISTEMAS DE ADQUISICIÓN DE IMÁGENES DE RESONANCIA MAGNÉTICA (MRI) SOBRE LOS PARÁMETROS ACÚSTICOS DE LA VOZ*

AUTOR: LUIS CARLOS ESPELETA CADENA**

PALABRAS CLAVES: MRI, Tracto vocal, señal de voz, USC-TIMIT, Fo, Jitter, shimmer, formantes, HNR, LPC, MFCC, LTAS, Matlab, Praat

DESCRIPCIÓN:

En los últimos años las *Magnetic Resonance Imaging* (MRI) se ha convertido en una poderosa herramienta usada en investigaciones sobre la producción del habla, puesto que permiten obtener imágenes 3D de la dinámica del tracto vocal en simultaneo con la señal de voz, lo cual proporciona una idea más clara sobre las relación acústico-articulatoria de la fonación. Sin embargo, esta técnica introduce un fuerte ruido acústico durante la obtención de las imágenes del tracto vocal y de la señal de audio. Por lo tanto, el presente trabajo se propone analizar y evaluar la influencia de sistemas de adquisición de imágenes de resonancia magnética sobre parámetros acústicos de la voz mediante la comparación entre la señal acústica con ruido MRI posfiltrada y señales sin ruido, ambas provenientes de la base de datos USC-TIMIT. Primero se extrajeron los siguientes parámetros descriptores de la voz : Frecuencia fundamental (Fo), jitter, shimmer,harmony to noise ration (HNR), los primeros cuatro formantes, lineal prediction coefficients (LPC), mel frequency cepstral coefficients (MFCC), y long term average spectrum (LTAS), tanto de las señales de audio limpia como de las señales de audio con ruido MRI, por medio de rutinas de programación escritas en los software Praat y Matlab, y finalmente ,los valores obtenidos fueron comparados con medidas de distancias, de calidad y

* Trabajo de grado

** Facultad de ingenierías Físico-Mecánicas. Ingeniería Electrónica. Director: Franklin Alexander Sepúlveda Sepúlveda ,PhD.

estadísticas. Los resultados obtenidos muestran que en general, la influencia del ruido producido por los sistemas de MRI es perceptible y los parámetros acústicos se ven afectados considerablemente. Los más afectados son el HNR y el primer formante.

ABSTRACT

TITLE: ANALYSIS OF THE INFLUENCE OF NOISE CAUSED BY SYSTEMS OF ACQUISITION OF MAGNETIC RESONANCE IMAGES (MRI) ON THE ACOUSTIC PARAMETERS OF THE VOICE*

AUTHOR: Luis Carlos Espeleta Cadena**

KEYWORDS: MRI, vocal tract, voice signal, USC-TIMIT, Fo, Jitter, shimmer, formants, HNR,LPC,MFCC,LTAS, Matlab, Praat

DESCRIPTION:

In recent years Magnetic Resonance Imaging (MRI) has become a powerful tool used in research on speech production, since they allow to obtain 3D images of vocal tract dynamics in simultaneous with the voice signal, which provides a clearer idea about the acoustic-articulatory relationship of phonation. However, this technique introduces a strong acoustic noise during the obtaining of the images of the vocal tract and the audio signal. Therefore, this work aims to analyze and evaluate the influence of magnetic resonance imaging systems on acoustic parameters of the voice by comparing the acoustic signal with post-filtered MRI noise and noise-free signals, both from the base of data USC-TIMIT. First, the following voice descriptor parameters were extracted: Fundamental frequency (Fo), jitter, shimmer, harmony to noise ratio (HNR), the first four formants, linear prediction coefficients (LPC), mel frequency cepstral coefficients (MFCC), and long term average spectrum (LTAS), both of the clean audio signals and of the audio signals with noise MRI, by means of programming routines written in the Praat and Matlab software, and finally, the values obtained were compared with measurements of distances, quality and statistics. The results obtained show that, in general, the influence of the noise produced by the MRI systems is perceptible and the acoustic

* Graduation project

** Physical- Mechanical Faculty. Electronic Engineering. Director: Franklin Alexander Sepulveda Sepulveda , PhD.

parameters are considerably affected. The most affected are the HNR and the first formant.

INTRODUCCIÓN

La resonancia magnética (MRI, *Magnetic Resonance Imaging*) es uno de los métodos fundamentales de diagnóstico clínico usados en hospitales y otros centros de atención médica¹. Durante la última década, se ha observado un aumento considerable en el uso de los escáneres de MRI, tanto en términos de nuevas aplicaciones clínicas, como en términos de mejoras tecnológicas, además del aumento de su disponibilidad en el mercado. Adicionalmente, la obtención de MRI es rápido, seguro, no invasivo y no emite radiaciones ionizantes²; lo cual ha provocado que sea utilizado, recientemente, para el análisis de la dinámica del tracto vocal.

Unas de las razones por la cual los sistema de MRI se han convertido en una poderosa herramienta usada para investigar la dinámica del tracto vocal es debido a que son capaces de obtener imágenes 3D del tracto vocal en simultaneo con la señal de voz, lo cual permite obtener una idea más clara de la forma y movimiento de los articuladores durante el proceso de la fonación.³ Esta relación entre la información contenida en la señal de voz y la posición de los órganos articuladores es conocida como: inversión acústica articulatoria y aunque ofrece una variedad de potenciales aplicaciones en el campo de la voz, es aún un problema por resolver⁴. Entre las posibles aplicaciones que se pueden implementar con un sistema que realice una adecuada reconstrucción de las configuraciones articulatorias se tienen: diseño de softwares que guían el aprendizaje de una segunda lengua, ayudas visuales para las terapias articulatorias en niños con dificultades auditivas y del habla; y en la mejora de los sistemas de reconocimiento del habla⁵.

¹ STEGGER, I, et al. Cardiac PET/MRI. *Curr Cardiovasc Imaging Rep.* Apr 2013, vol. 6, no. 2, pp. 169

² *Ibíd.*, p. 169.

³ NARAYANAN, S, et al. An approach to real-time magnetic resonance imaging for speech production. *Acoustical Society of America.* 2004, p. 1772.

⁴ SEPULVEDA, A; CASAS, D y CASTELLANOS, G. Importancia de las frecuencias de resonancia del tracto vocal en la estimación de posiciones articulatorias. *Revista Ingeniería Biomédica.* 2012, vol. 6, nº 11, p. 57.

⁵ *Ibíd.*, p. 57

Sin embargo, el uso de los sistemas de MRI para la investigación de la inversión acústica-articulatoria presenta la desventaja que durante el proceso de adquisición de las imágenes y del audio se genera un fuerte ruido acústico. La vibración provocada por las fuerzas de Lorentz en un escáner de MRI no solo se propaga a otras partes del escáner y la estructura circundante, sino que también induce ruido acústico⁶ y, su efecto depende de la intensidad del campo magnético. Con el fin de conseguir imágenes con mayor resolución se requieren campos magnéticos fuertes y/o tiempos de escaneo más prolongados. Para reducir el tiempo de escaneo, y así poder obtener videos de resonancia magnética se requiere un campo magnético más fuerte y, por lo tanto, una corriente de gradiente más alta, lo que produce mayores niveles de ruido sobre la señal de voz. Este es el caso de escáneres modernos, los cuales emplean corrientes de conmutación más rápidas y tiempos de escaneo más cortos bajo fuertes campos magnéticos, por lo tanto, la vibración resultante y el ruido acústico asociado son aún mayores. Este ruido acústico muestra un ancho de banda aproximadamente de 8 kHz con un nivel de presión sonora entre 123 y 145 dB⁷.

En tareas de tipo multimodal en las que se busca analizar la señal acústica en concurrencia con los videos MRI, es necesario utilizar estrategias que disminuyan ostensiblemente el nivel de ruido sobre la señal acústica. Al incluir un micrófono dentro de la cabina, el metal presente en la mayoría de micrófonos causaría un bajo rendimiento en el centro del imán e interferiría con la secuencia de escaneo, distorsionando así las imágenes finales⁸. Igualmente, el ruido de la secuencia de escaneo es bastante alto y sería difícil separar la voz de los pacientes. Por lo tanto, en este tipo de tareas se plantea el uso de micrófonos de fibra óptica (MFO), e.g. (Phone-or corporation, Or-Yehuda, Israel)*, el cual incorpora mecanismos de

⁶ KHAN, I. MRI Scanner's Vibration Isolation: Experimental Measurements, Analysis Techniques and Analytical Models. 31 January 2017, p. 351.

⁷ MONTAZERI, V; PATHAK, N y PANAHI, I. Two-channel multi-stage speech enhancement for noisy fMRI environment. *IEEE.2013*, vol. 36, nº 2, p. 62.

⁸ NESSAIVER, M, *et al.* Recording High Quality Speech During Tagged Cine-MRI Studies Using a Fiber Optic Microphone. *Journal of Magnetic Resonance Imaging*. 2006, p. 92.

* <http://phone-or.com>

cancelación de ruido. Aún así, se requiere que la voz captada por los micrófonos sea luego filtrada mediante algoritmos basados en filtros adaptativos que tienen en cuenta la estructura periódica del ruido generado, tal como se realizó durante la recolección de la base de datos USC-TIMIT⁹.

La gran mayoría de trabajos previos relacionados con la adquisición de la señal de voz por medio de sistemas de MRI, se han centrado en la evaluación de los algoritmos de filtrado diseñados para reducir los niveles del ruido^{10 11 12}. El rendimiento de sus investigaciones lo miden de acuerdo a la calidad de la voz obtenida después de aplicar el filtro propuesto o los filtros propuestos. Esta calidad la determinan de acuerdo a metodologías de pruebas subjetivas y objetivas que evalúan los sistemas de comunicación vocal que utilizan un algoritmo de cancelación de ruido, por ejemplo, la norma ITU-T - P 8.35¹³ o la medida PESQ¹⁴.

Por lo tanto, el presente trabajo de grado, además de tener en cuenta la calidad de la señal de voz grabada con los sistemas de adquisición de imágenes de resonancia magnética, tiene por objetivo principal analizar y evaluar la influencia de estos sistemas sobre las medidas acústicas que proporcionan información relevante de la voz, mediante la comparación entre la señal acústica con ruido MRI posfiltrada y señales sin ruido de MRI, ambas provenientes de la base de datos USC-TIMIT. Esta base de datos incluye videos de resonancia magnética de cinco hombres y cinco mujeres hablantes del inglés estadounidense; además, incluye registros de audio de las mismas personas pronunciando las mismas frases sin la

⁹ NARAYANAN, S, et al. USC-TIMIT: A database of multimodal speech production data. Acoustical Society of America, 2014, pp. 11.

¹⁰ CHATAN,V; CHANDAN,R e ISSA,P. Improving Quality and Intelligibility of Speech using Single Microphone for the Broadband fMRI Noise at Low SNR. IEEE. 2016, pp. 3675.

¹¹ PRIBIL, J; PRIBILOVA, A y FROLLO, I. Vibration and Noise in Magnetic Resonance Imaging of the Vocal Tract: Differences between Whole-Body and Open-Air Devices. Sensors.5 Abril 2018, pp. 351

¹² PRIBIL,J; PRIBILOVA,A y FROLLO, I. Automatic GMM-Based evaluation of noise suppression in the speech signal recorded during phonation in the open-air MRI. De 2017 40th International Conference on Telecommunications and Signal Processing (TSP).Barcelona, 2017,vol. 17, nº 6, pp. 257

¹³ I.-T. R. p.835. Subjective Test Methodology for Evaluating Speech Communication Systems that Include Noise Suppression Algorithm. 2003.

¹⁴ ITU_T Recommendation, P.862, 2001.Perceptual Evaluation of Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs, 2001.

influencia de efectos causados por el sistema MRI, pero tomadas en una sesión diferente (datos de articulografía electromagnética, EMA, *Electromagnetic Articulograph*). Para la reducción del ruido acústico, esta base de datos utiliza un algoritmo de procesamiento de señal adaptativo que tiene en cuenta la estructura periódica del ruido generado por el escáner de adquisición de MRI.

Para lograr el objetivo principal del presente trabajo, primero se propone extraer medidas acústicas de la señal de voz, tanto de la señal de audio afectada por el sistema MRI como de la señal sin ruido (EMA). Luego, estas medidas son comparadas con el uso de medidas de distancias espectrales y medidas estadísticas. Adicionalmente se utilizan medidas objetivas de la calidad de la voz para medir el grado de afectación.

1. OBJETIVOS Y ALCANCE:

1.1 OBJETIVOS GENERALES

Analizar y evaluar la influencia de sistemas de adquisición de imágenes de resonancia magnética (MRI) sobre parámetros acústicos de la voz mediante la comparación entre la señal acústica con ruido MRI posfiltrada y señal sin ruido provenientes ambas de la base de datos USC-TIMIT.

1.2 OBJETIVOS ESPECÍFICOS

- Extraer medidas acústicas relacionadas con la periodicidad y la envolvente espectral de la señal de voz, tanto de la señal de audio filtrada como sin filtrar.
- Evaluar el grado de influencia del sistema de adquisición de MRI y filtrado acústico sobre señales de voz mediante el uso de medidas de comparación de distancias y medidas de tipo estadístico.

1.3 ALCANCE

El presente trabajo se propone evaluar el grado de influencia del sistema de adquisición de MRI con etapa de postfiltrado sobre la señal acústica de voz. En particular, se analiza la influencia sobre aquellos parámetros descriptores de la voz más relevantes. Dentro de las medidas acústicas se da preferencia a aquellas pertenecientes a la categoría de medidas relacionadas con la periodicidad y la envolvente espectral de la señal de voz. El análisis de la influencia del ruido de MRI se realiza sobre señales de voz de habla continua, en el idioma inglés pertenecientes bases de datos disponibles. En particular la base de datos USC-TIMIT contiene videos de resonancia magnética (RM) de la sección sagital del tracto vocal en paralelo con registros de audio.

2. MARCO TEÓRICO

2.1 PRINCIPIOS BÁSICOS DE LOS SISTEMAS DE MRI

MRI es una técnica espectroscópica usada en el campo médico para producir imágenes del interior del cuerpo humano¹⁵. Está basada en los principios de la resonancia magnética nuclear (RMN), la cual es una técnica espectroscópica usada para obtener datos físicos y químicos microscópicos de las moléculas^{16 17}. También hay que agregar que la base de las MRI es el campo magnético direccional, o momento, asociado con las partículas cargadas en movimiento. Los núcleos que contienen un número impar de protones y/o neutrones realizan un movimiento característico o precesión. Debido a que los núcleos son partículas cargadas, esta precesión produce un pequeño momento magnético¹⁸. Cuando un cuerpo humano es inmerso en un campo magnético elevado, algunos de los núcleos de hidrógeno libres se alinean con la dirección del campo magnético. Este comportamiento se denomina precesión de Larmor¹⁹.

La frecuencia de la precesión de Larmor es proporcional a la intensidad del campo magnético aplicado, y definida como:

$$\omega_0 = \gamma \beta_0, \quad (1)$$

donde γ es razón giromagnética y β_0 es la intensidad del campo aplicado. La constante giromagnética es una constante específica del núcleo de los elementos. Para el hidrógeno, $\gamma=42.6$ MHz/Tesla²⁰.

¹⁵ STEGGER. Op. cit., p. 171

¹⁶ CHATAN, V; CHANDAN, R e ISSA, P. Op. cit., p. 3676

¹⁷ SHARMA, H. MRI physics-basic principles. Acta Neuropsychiatrica. Agosto 2009, p. 200.

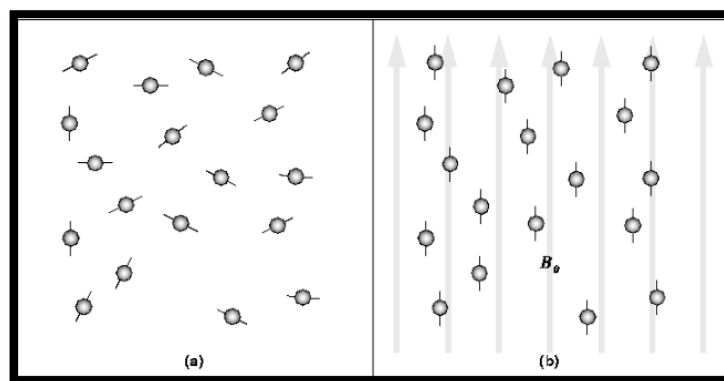
¹⁸ PRIBIL, J; PRIBILOVA, A y FROLLO, I. Op. cit., p. 258

¹⁹ SCOTT, A. Speech MRI: Morphology and function. Physica Medica. 2014, vol. 30, p. 604.

²⁰ MACKIEWICH, B. A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science. Intracranial Boundary Detection and Radio Frequency Correction in Magnetic Resonance Images, Simon Fraser University, 1995.

Así, para obtener una MRI de un objeto, el objeto debe ser ubicado en un campo magnético uniforme, β_0 , de 0.5 a 1.5 Tesla²¹. Como resultado, los núcleos de hidrógeno del objeto se alinean con el campo magnético y crean un momento magnético neto M paralelo a β_0 . Este comportamiento se ilustra en la Figura 1. En ausencia de un campo magnético fuerte, los núcleos de hidrógeno están alineados aleatoriamente como en (a). Cuando un campo magnético fuerte, β_0 , es aplicado, los núcleos de hidrógeno precesan sobre la dirección del campo como en (b).

Figura 1. Núcleos de Hidrogeno en ausencia y presencia de un campo magnético fuerte.

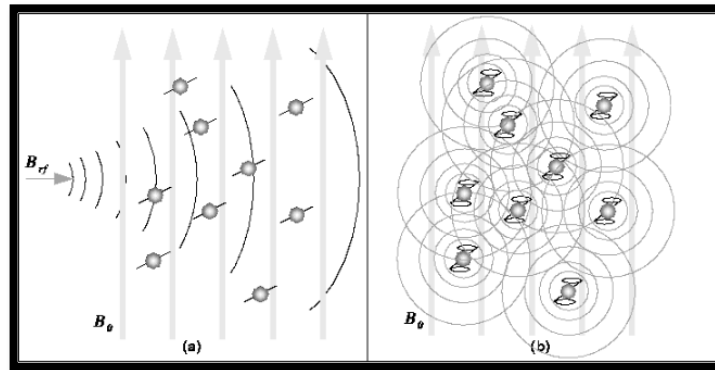


Fuente: MACKIEWICH, B.

Después, un pulso de radio frecuencia (RF), β_{rf} , es aplicado perpendicularmente a β_0 . Este pulso, con una frecuencia igual a la frecuencia de Larmor, causa que M , se incline en contra de β_0 como en la Figura 2a.

²¹ CHATAN, V; CHANDAN, R e ISSA, P. Op. cit., p. 3674.

Figura 2. (a) Pulso RF aplicado. (b) Cuando se detiene el pulso RF.



Fuente: MACKIEWICH, B.

Una vez que la señal RF es removida, los núcleos se realinean, de manera que su momento magnético neto M , es nuevamente paralelo con B_0 . Este retorno al equilibrio es denominado como *relajación*. Durante la relajación, los núcleos pierden energía y emiten su propia señal RF (ver Figura 2b). Esta señal es denominada como la señal de respuesta de inducción de caída libre (ICL). La señal de respuesta ICL es medida por una bobina de campo conductivo ubicada alrededor del objeto que se está reconstruyendo. Esta medida es procesada para obtener una MRI 3D en escala de grises.

2.2 COMPONENTES DE UN SISTEMA DE MRI

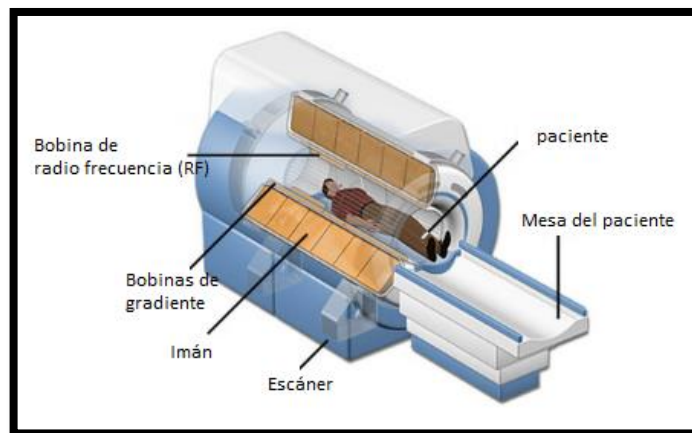
Las máquinas de MRI varían tanto en tamaño como en forma. Los diseños más antiguos contaban con un espacio más reducido y compacto y eran muy cerradas. Esto último afectaba a los pacientes mentalmente, y a menudo provocaba en ellos miedo, incluso antes de ser examinados²². Sin embargo, los ingenieros han tratado de resolver este problema al lograr que las máquinas sean más abiertas y acogedoras. Ellos han expandido sus lados e incluyeron más espacio en el escáner

²² PRIBIL, J; PRIBILOVA, A y FROLLO, I. Op. cit., p. 351.

que el que había en los modelos originales. El diseño básico de un escáner de MRI es el mismo en casi todas las máquinas.

El escáner consiste en un tubo de aproximadamente 0.6 m de diámetro, dentro del cual se lleva a cabo el examen a los pacientes. Además, contiene un imán, una bobina de radio frecuencia (RF), tres bobinas de gradientes, la mesa del paciente, una bomba criogénica y un sistema de cómputo. En la Figura 3 se muestra un escáner típico de MRI y sus componentes:

Figura 3. : Escáner típico de MRI y sus componentes.



Fuente: RASHID, N; KIBRIA, F.

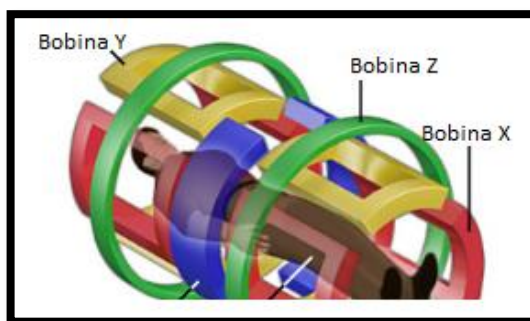
- **Imán superconductor:** Es la parte más importante y grande de un dispositivo de MRI. También es el que le permite a la máquina de MRI producir imágenes de alta calidad. Hay que agregar que el imán es extremadamente poderoso y su intensidad es medida en Tesla o Gauss (1 Tesla=10000 Gauss)²³. La mayoría usan un campo magnético de 0.5 a 2.0 tesla, cuando el campo magnético de la Tierra es de solo 0.5 gauss. Se enfrían a la temperatura de los superconductores (-269° C) usando helio o

²³ RASHID, N; KIBRIA, F. MRI SCAN. Weebly. [En línea]. Disponible en: <https://snc2dmri.weebly.com/components--functions.html>.

nitrógeno, por lo tanto, la corriente eléctrica que fluye por él lo hace de forma permanente, creando un campo magnético constante.

- **Bomba Criogénica:** Es la encargada de llenar constantemente el helio que se usa para enfriar el imán
- **Bobinas de gradiente:** Hay tres diferentes bobinas de gradiente que están dentro de la máquina de MRI y están localizadas dentro del imán principal. Cada una de esas bobinas produce un campo magnético menos intenso que el campo principal (ver Figura 4). Asimismo, las bobinas de gradiente crean un campo variable (x, y, z) que pueden crecer o decrecer al alterar y ajustar el campo magnético principal y así escanear partes del cuerpo específicas²⁴.

Figura 4. Las tres bobinas de gradiente.



Fuente: RASHID, N; KIBRIA, F.

- **Bobinas de Radio Frecuencia (RF):** La función básica de las bobinas de RF es transmitir las ondas de radio frecuencia dentro del cuerpo del paciente (ver Figura 5). Existen bobinas localizadas dentro del escáner de MRI para transmitir ondas dentro de diferentes partes del cuerpo. De igual manera, si se especifica una determinada área del cuerpo, entonces todas las bobinas

²⁴ SHARMA, H. Op. cit., p. 200.

de RF generalmente se enfocan en la parte del cuerpo que se está escaneando para permitir una mejor exploración.

Figura 5. Bobina de Radio Frecuencia (RF).



Fuente: RASHID, N; KIBRIA, F. MRI SCAN. Weebly. [En línea]. Disponible en: <https://snc2dmri.weebly.com/components--functions.html>.

- **Mesa del paciente:** Este componente simplemente desliza al paciente dentro de la máquina de MRI. La posición en la que el paciente se acuesta sobre la mesa está determinada por la parte del cuerpo a escanear. Una vez que la parte de cuerpo que se está examinando está en el centro exacto del campo magnético, el cual es conocido como isocentro, el proceso de escaneo inicia²⁵.
- **Antena/ sistema de cómputo:** La antena es un dispositivo muy sensible que fácilmente detecta las señales de RF emitidas por el cuerpo del paciente mientras es analizado y envía esta información al sistema de cómputo, cuya principal función es recibir, grabar y analizar las imágenes del cuerpo del paciente que ha sido escaneado. Adicionalmente, este sistema interpreta los

²⁵ MECHEFSKE, C, et al. Acoustic noise reduction in a 4 T MRI scanner. MAGMA. 2002, vol. 13, pp. 173.

datos enviados por la antena y luego ayuda a producir una imagen comprensible de la parte del cuerpo examinada.

Los datos trabajados en el presente trabajo de grado fueron adquiridos en el Hospital County de Los Ángeles en un escáner Signa Excite HD 1.5 T (GE Healthcare, Wausesha WI) con gradientes capaces de generar 40mT/m de amplitud y 150 mT/m/ms de velocidad de giro²⁶.

2.3 CARACTERÍSTICAS DEL RUIDO ACÚSTICO DE MRI

El ruido acústico de MRI se genera durante el escaneo al conmutar rápidamente los gradientes de campo sujetos a placas de soporte para alternar las fuerzas de Lorentz transversales²⁷. Asimismo, el generador de MRI escanea un número fijo de secciones transversales a lo largo del plano transversal de las imágenes a obtener y envía pulsos de radio frecuencia (RF) modulados periódicamente. Los pulsos de RF son generados al enviar pulsos de corriente a través del gradiente de una bobina, por lo tanto, el material de soporte es sometido a una fuerza de Lorentz periódica, la cual causa una vibración estructural. Lo anterior produce un ruido acústico cuasi-periódico de banda ancha con altos niveles de energía, típicamente en un rango de frecuencias de hasta 8 KHz y con niveles de presión sonora entre 123 y 145 dB²⁸.

La Figura 6 ilustra los dispositivos presentes en una sala de escaneo de MRI, la localización de los micrófonos y las principales fuentes de ruido presentes en tal sala²⁹. Las principales fuentes de ruido son: el paciente en sí mismo, el escáner MRI y la bomba criogénica. La línea punteada simboliza la trayectoria del sonido. Al omitir las reflexiones sobre las paredes de la habitación se tiene que: la voz del

²⁶ NARAYANAN, S, et al. Op. cit., p. 11.

²⁷ MONTAZERI, V; PATHAK, N y PANAHI, I. Op. cit., p. 63.

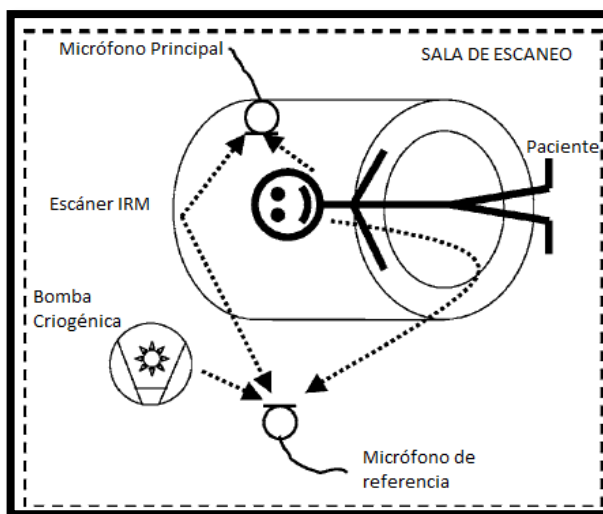
²⁸ *Ibíd.*, p. 63.

²⁹ BRESCH, E, et al. Op. cit., p. 1793.

paciente es primero que todo captada por el micrófono principal, sin embargo, hay un camino de fuga hacia el micrófono de referencia³⁰. Por consiguiente, el gradiente de ruido MRI es captado por los dos micrófonos a través de diferentes trayectorias, y por lo tanto, con diferentes tiempos de retardo, pero con una intensidad similar. Por último, el ruido debido a la bomba criogénica afecta principalmente al canal de referencia.

De otra parte, el escáner de MRI también tiene integrado un ventilador de enfriamiento, el cual produce un flujo de aire a través de la abertura del imán. Por consiguiente, el ventilador puede producir un ruido adicional, pero puede ser apagado durante el escaneo. Se debe notar que el gradiente de ruido MRI, es por lejos el más fuerte de todas las fuentes de ruido. Y a pesar de su alta potencia, el ruido MRI también tiene algunas características ventajosas, por ejemplo, es estacionario, periódico y depende directamente de la secuencia de pulso MRI³¹.

Figura 6. Fuentes de ruido y ubicación de los micrófonos en una sala de escaneo de MRI.



Fuente: BRESCH, E, et al.

³⁰ BRESCH, E, et al. Op. cit., p. 1794.

³¹ MONTAZERI, V; PATHAK, N y PANAHI, I. Op. cit., p. 64.

A continuación, se da una descripción más detallada de las distintas fuentes de ruido:

Bomba criogénica: Se usa para mantener el helio o el nitrógeno en estado líquido. El helio evaporado es reciclado por compresión y luego expandido para ser licuado nuevamente. Esta bomba produce un fuerte sonido, principalmente a bajas frecuencias. El nivel de presión del sonido producido alcanza hasta los 80 dB³². Si se apaga la bomba el helio comienza a ebulir; y si se apaga por un periodo prolongado puede provocar que las bobinas magnéticas pierdan su superconductividad debido al aumento de la temperatura.

Cables de corriente eléctrica: Unos cables de corriente débilmente ajustados a las bobinas de gradiente pueden producir un fuerte ruido a la frecuencia de resonancia de la longitud del cable. La corriente alterna a través de los cables interactúa con el campo magnético estático; así, las fuerzas de Lorentz evocadas golpean los cables al criostato. Estas fuentes de ruido se pueden evitar al ajustar las conexiones de los cables o al colocar cables de corriente rígidos a cierta distancia de otros objetos³³.

Corrientes de Foucault: Las corrientes de Foucault pueden ser inducidas en las bobinas de RF, en el escudo de RF y en el campo magnético estático de la bomba criogénica³⁴. Campos magnéticos variables, como los campos de gradiente en MRI, producen un campo eléctrico variable de acuerdo las leyes de Maxwell. Además, la ley de Lenz establece que un campo magnético variable induce un campo magnético opuesto en un conductor³⁵. Por esta razón, las corrientes de gradiente eléctrico tienen un rebasamiento en el comienzo de un pulso para compensar este efecto. Además, las corrientes de Foucault inducidas son variantes en el tiempo y la fuerza de Lorentz resultante hace que el conductor vibre. Como las bobinas de RF están ubicadas cerca del oído del paciente, estas vibraciones pueden ser una

³² HOITING, G. Measuring MRI noise. Groninga, 2005

³³ RONDINONI, C. Effect of scanner acoustic background noise on strict resting-state fMRI. Brazilian Journal of Medical and Biological Research. 2013, p. 360.

³⁴ SCOTT, A. Op. cit., p. 606.

³⁵ KHAN, I. Op. cit., p. 352.

fuentes importantes de ruido. Igualmente, las corrientes de Foucault en el escudo de RF y en la bomba criogénica, hacen que estas estructuras vibren y resulte un ruido acústico³⁶.

Ruido de las bobinas de gradiente: Las fuertes corrientes eléctricas que fluyen a través de las bobinas de gradiente interactúan con el campo magnético estático. Las fuerzas de Lorentz resultantes se distribuyen sobre las bobinas de gradiente e intentan desviar la estructura de las bobinas de gradiente. Los patrones de vibración reales dependen de la distribución de la fuerza, las propiedades del material, las frecuencias de excitación y las dimensiones de las bobinas de gradiente. Estas vibraciones pueden coincidir con las frecuencias propias del ensamblaje de las bobinas de gradiente, lo cual resulta en fuertes vibraciones y altos niveles de presión sonora³⁷.

³⁶ MECHEFSKE, C, *et al.* Op. cit., p. 174.

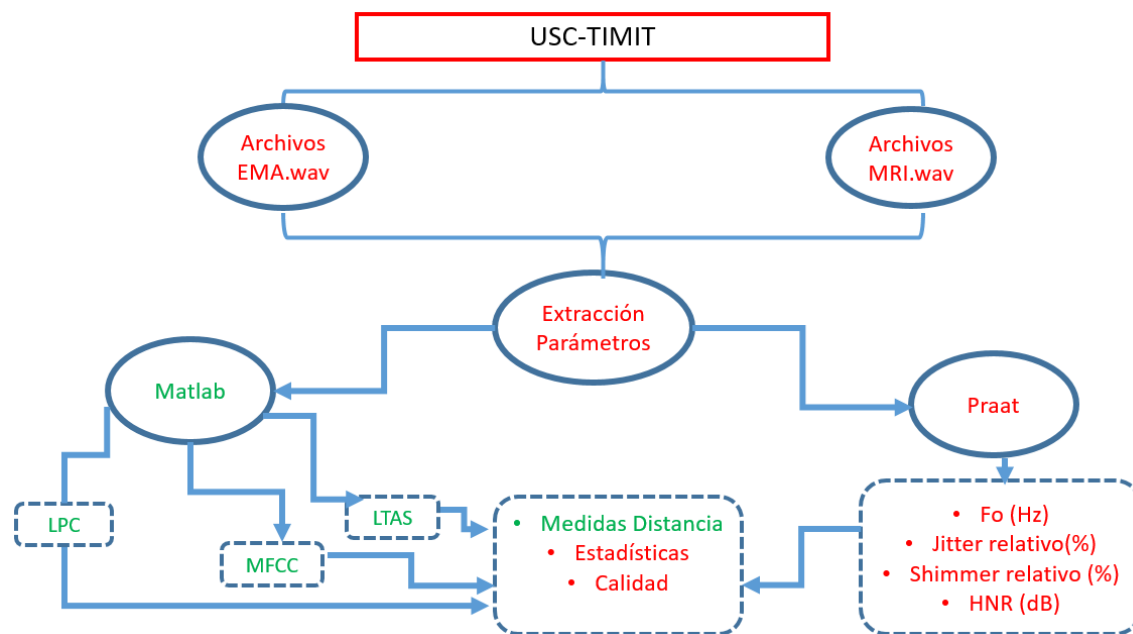
³⁷ SCOTT, A. Op. cit., p. 604.

3. MÉTODO

La base de datos USC-TIMIT está organizada de la siguiente manera: cada uno de los 10 hablantes cuenta con 92 archivos de audio.wav de MRI. Y cada archivo de audio se compone de cinco frases pronunciadas por el hablante. Por lo tanto se tiene un total de 460 frases (ver ANEXO A) de MRI por hablante. Esta misma organización aplica para los archivos de audio.wav de EMA, con la diferencia que estos solo están disponibles para 5 de los 10 hablantes de la base de datos. En la sección 2.1 se abordará con más detalle la base de datos USC-TIMIT.

Por otra parte, el método seguido para lograr los objetivos propuestos se muestra en la Figura 7:

Figura 7. Método seguido durante la realización del proyecto de grado



- Primero se realiza una normalización en el tiempo de las señales de voz limpias (archivos EMA.wav) y de las señales de voz con ruido (archivos MRI.wav). Esta normalización consistió en remover la media de las señales de voz y dividir entre su desviación estándar, con el fin de remover niveles de DC. También se eliminaron los segmentos de silencios entre frase y frase,

tanto de los registros de audio limpios como con ruido. Para ellos se usó la función.

- Segundo, se extrajeron parámetros o medidas acústicas (sección 3) que brindan información relevante sobre la señal de voz, se hizo tanto para las frases de los archivos EMA como de los archivos MRI.
- Las medidas acústicas de las señales de voz limpias y con ruido se compararon por medio de medidas de distancia y medidas estadísticas (sección 4). También se evaluó la calidad de la voz adquirida por sistema de MRI, por medio de medidas objetivas de calidad (sección 5) desarrolladas a partir de metodologías de pruebas subjetivas que evalúan los sistemas de comunicación vocal que utilizan un algoritmo de cancelación de ruido

3.1. BASE DE DATOS USC-TIMIT

Los datos de MRI en tiempo real (MRItr) han sido adquiridos de diez hablantes nativos del Inglés Americano, ninguno de los cuales reportó anomalías auditivas ni del habla, con características indicadas en la Tabla 1, en donde, la última columna indica si los datos de articulografía electromagnética* están disponibles. En total, la base de datos USC-TIMIT cuenta con datos de articulografía electromagnética (EMA, *Electromagnetic Articulograph*) para cinco de esos hablantes. La Figura 8 muestra algunos ejemplos de imágenes del tracto vocal de esos hablantes. Estas dos modalidades, MRI y EMA, fueron registradas en dos sesiones independientes mientras los sujetos producían las mismas 460 frases pertenecientes a la base de datos MOCHA-TIMIT³⁸. El conjunto de oraciones está diseñado para obtener todos los fonemas del inglés americano, dentro de un amplio rango de contextos prosódicos y fonológicos.

* En la siguiente dirección web se puede encontrar las especificaciones de los articulógrafos electromagnéticos usados: <http://www.articulograph.de/>

³⁸ NARAYANAN, S, et al. Op. cit., p. 10.

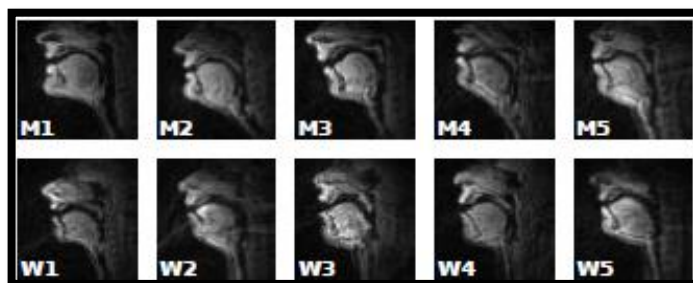
Tabla 1. Participantes del estudio-Detalles demográficos.

ID	GÉNERO	EDAD	LUGAR DE NACIMIENTO	EMA
M1	Masculino	29	Buffalo,NY	✓
M2	Masculino	33	Ann Arbor,MI	✓
M3	Masculino	26	Madison,WI	
M4	Masculino	26	St. Louis, MO	✓
M5	Masculino	27	Mammoth,CA	
W1	Femenino	23	Commack,NY	✓
W2	Femenino	32	Westfield,IN	
W3	Femenino	20	Palos Verdes,CA	
W4	Femenino	46	Pittsburgh,PA	
W5	Femenino	25	Brawley,CA	✓

Fuente: NARAYANAN, S, et al.

Además de proporcionar una muestra fonológica completa del inglés, este corpus fue elegido para brindar una herramienta adicional para los investigadores que han utilizado previamente la base de datos MOCHA-TIMIT. La base de datos y las técnicas complementarias para la reconstrucción, procesamiento y análisis lingüístico están disponibles para la comunidad científica.

Figura 8. Tractos vocales de los diez hablantes.



Fuente: NARAYANAN, S, et al.

3.1.1 Adquisición de la señal de audio. Las señales de audio fueron grabadas, al interior del escáner de MRI, en paralelo con los videos MRI usando una frecuencia de muestreo de 20 KHz. Para ello se utilizaron micrófonos de fibra-óptica (*Phone-or corporation, Or-Yehuda, Israel*). La sincronización con la señal de video se controló mediante el uso de un reloj de muestreo derivado del reloj maestro de 10 MHz del escáner.

En la Figura 9 se muestran los diferentes componentes del sistema de adquisición de datos y su localización en la sala de escaneo, en la sala de sistemas y en la sala de control. Se pueden observar los dos micrófonos de fibra óptica localizados en la sala de escaneo. El micrófono principal está aproximadamente a una distancia de 1.3 cm (0.5 in) de la boca del paciente y en un ángulo de 20 grados. El micrófono de referencia está ubicado por fuera del escáner, aproximadamente a una distancia de 0.9 m (3ft) de la pared lateral y a una altura alrededor de los 1.2 m (4ft). Los micrófonos están conectados a la caja óptica receptora, la cual está ubicada en la sala de control de MRI. Los datos son grabados en un computador laptop usando la tarjeta NI-DAQ 6036E PCMCIA de *National Instruments*^{*}, la cual proporciona una tasa de muestreo de 200 kHz y admite hasta 16 canales de entradas analógicas³⁹.

^{*} En la siguiente dirección web se puede encontrar las especificaciones de tarjeta NI-DAQ 6036E PCMCIA: <http://www.ni.com/es-co/support/model.daqcard-6036.html>

³⁹ NARAYANAN, S, et al. Op. cit., p. 13.

Figura 9. Diagrama de adquisición de audio de la base de datos USC-TIMIT.

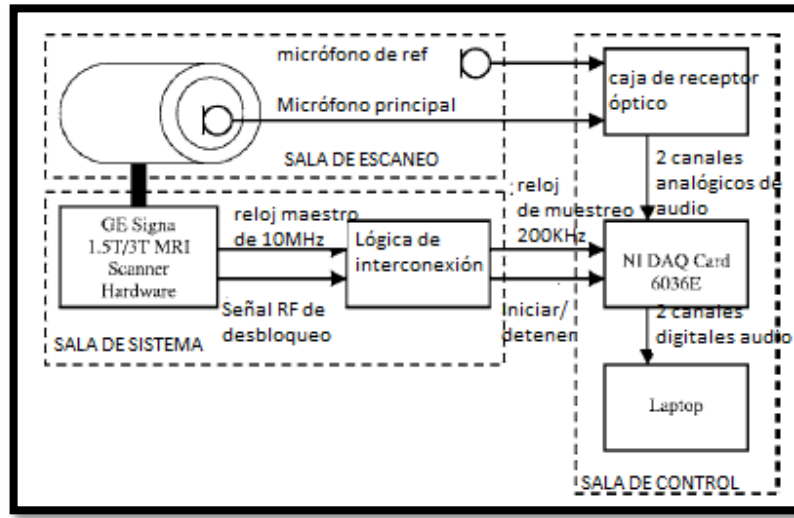
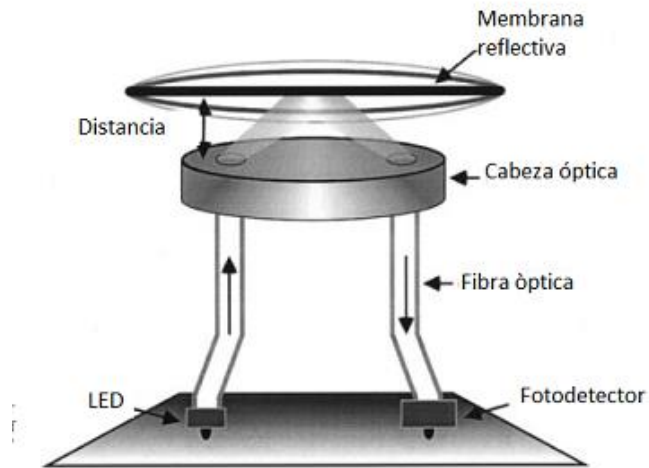


Figura. 1: NARAYANAN, S, et al.

Micrófono de fibra óptica (MFO): El principio de operación del MFO es soportado por un sensor óptico que mide la distancia a una membrana al detectar la intensidad de la luz que es reflejada en la membrana⁴⁰. Como se puede ver en la Figura 10, la luz emitida por un diodo emisor de luz (LED) recorre una fibra óptica hasta la cabeza óptica, que a su vez envía la luz a una membrana sensible al sonido. El sonido hace que la membrana vibre, cambiando así la intensidad de la luz reflejada en la membrana. Esta luz se transmite de vuelta al fotodetector a través de una segunda fibra óptica, que transforma la luz modulada en intensidad en una señal de audio por medio de un simple procesamiento electrónico.

⁴⁰ NESSAIVER, M, et al. Op. cit., p. 94.

Figura 10. Principio de operación de los MFO.



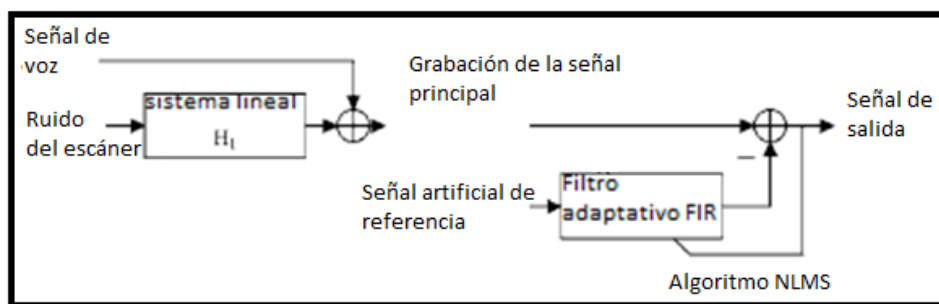
Fuente: MONTAZERI,V; PATHAK,N y PANAHI,I.

3.1.2 Reducción del ruido acústico en la base de datos USCM-TIMIT. El ruido generado a partir de la operación de los escáneres de MRI debe ser reducido satisfactoriamente con el fin de realizar posteriores análisis e investigaciones sobre la señal de voz. La solución usada consiste primero en mitigar el ruido de MRI al grabar la señal de voz con el uso de micrófonos de fibra óptica. Luego, la señal de voz puede ser filtrada por medio de algoritmos de filtros adaptativos que tengan en cuenta las propiedades del ruido.

El filtro usado fue desarrollado basado en el conocido algoritmo de Normalized Least Mean Square (NLMS). Usa una señal artificial de referencia, generada en base a un modelo del gradiente de ruido de MRI. El correspondiente diagrama del sistema se muestra en la Figura 11. El orden del filtro es de 4000, con una actualización de coeficientes de 0.5. Por otra parte, la desventaja de este método es que no tiene en cuenta otras fuentes de ruido presentes durante el escaneo de MRI, como, por ejemplo, la bomba criogénica.

Sin embargo, la mayor ventaja de este método es que no existen fugas de la señal deseada, es decir, la voz del sujeto, dentro del canal de referencia.

Figura 11. Método de cancelación del ruido usado en la base de datos USCM-TIMIT.



Fuente: BRESCH, E, et al.

3.2 REPRESENTACIÓN DE LA SEÑAL DE VOZ

3.2.1 Formantes. Corresponden a las “frecuencias más relevantes de resonancia del tracto vocal”⁴¹. También se definen como “las frecuencias naturales o propias de la trayectoria del aire en el tracto vocal desde la glotis hasta los labios”⁴². En ⁴³ analizan la contribución de los formantes en la inferencia de las posiciones de los articuladores del tracto vocal durante el proceso de la fonación.

Se escribió una rutina en el software para el análisis de la voz Praat* para hallar de forma automática el valor de los primeros 4 formantes de cada una de las frases pronunciadas por los hablantes de la base de datos. En la figura 12 se observan los parámetros solicitados por PRAAT y sus correspondientes valores para el cálculo de los formantes, estos son los que por defecto usa el software cuando se analiza una voz femenina; en el caso de una voz masculina, solo cambia valor del parámetro Maximun formant (Hz): el cual es de 5000 Hz.

Los resultados de este experimento se encuentran en el apartado de: Resultados y discusiones.

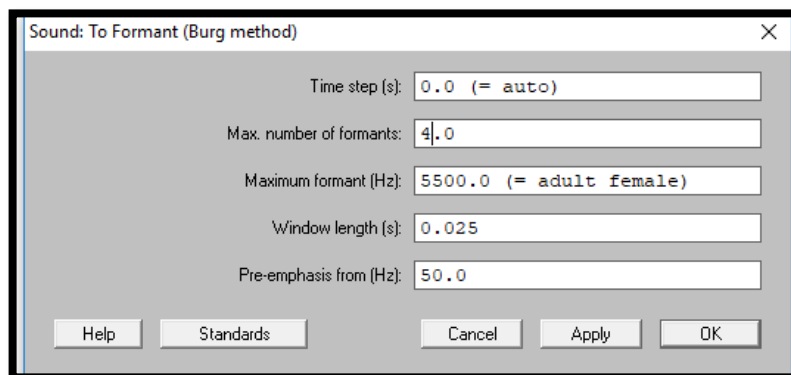
⁴¹ SEPULVEDA, A; CASAS, D y CASTELLANOS, G. Op. cit., p. 55.

⁴² RAMANARAYANAN, V, et al. Analysis of Speech Production Real-Time MRI. Computer Speech & Language. 2018, p. 221.

⁴³ SEPULVEDA, Alexander. Estimation of Articulatory Parameters from the Acoustic Speech Signal. Manizales, 2012, p. 13.

* En la siguiente dirección web se puede descargar el software Praat : <http://www.fon.hum.uva.nl/praat/>

Figura 12. Parámetros usados para el cálculo de los formantes con el software PRAAT.



3.2.2 Frecuencia Fundamental (Fo). El pitch o frecuencia fundamental F_0 se determina por la velocidad de apertura o cierre de las cuerdas vocales en la laringe durante la fonación de sonidos del tipo sonoro⁴⁴. Se halló la frecuencia fundamental de cada frase pronunciada por los hablantes de la base de datos por medio del software para el análisis de la voz Praat. En la figura.13 se observan los parámetros solicitados por Praat y sus correspondientes valores, estos son los que por defecto usa el software. Para el cálculo del pitch, Praat aplica un algoritmo que funciona con el método de autocorrelación, es decir, hace una búsqueda de los valores de la señal en una ventana de análisis hasta encontrar un patrón. En⁴⁵ es descrito este algoritmo.

⁴⁴ SEPULVEDA, Alexander; CASTELLANOS, G. Estimación de la frecuencia fundamental de señales de voz usando transformada Wavelet. Scientia et Technica.2004, nº 24, p. 8.

⁴⁵ BOERSMA, P. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound.Proceedings of the Institute of Phonetic Sciences. 1993, nº 17, p. 99.

Figura 13. Parámetros usados para el cálculo de Frecuencia fundamental con el software PRAAT.

Los resultados de este experimento se encuentran en el apartado de: Resultados y discusiones.

3.2.3 Jitter (relativo). Es la variación ciclo a ciclo de la frecuencia fundamental, es decir, el promedio de la diferencia absoluta entre periodos consecutivos, dividida entre el periodo promedio. Se expresa en porcentaje⁴⁶. Ha sido usado básicamente en estudios que tratan sobre patologías de la voz⁴⁷ y en sistemas de reconocimientos del hablante⁴⁸.

$$Jitter(relativo) = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i+1}|}{\frac{1}{N} \sum_{i=1}^N T_i}, \quad (2)$$

⁴⁶ FARRÚS, M; HERNANDEZ, J y EJARQUE, P. Jitter and shimmer measurements for speaker recognition. In proceedings of 8th annual conference of the international speech communication association (InterSpeech).2007, p. 780.

⁴⁷ ALÍAS, F; SOCORÓ, J y SEVILLANO, X. A review of physical and perceptual feature extraction techniques for speech, music and enviromental sounds. Applied Sciences, 2016, p. 351.

⁴⁸ FARRÚS, M; HERNANDEZ, J y EJARQUE, P. Op. cit., p. 780.

donde T_i son las longitudes extraídas del periodo F_0 y N es el número de periodos extraídos.

Se usó una rutina de programación sobre el software Praat para calcular de forma automática el jitter (relativo) de cada una de las frases pronunciadas por los hablantes. Los resultados de este experimento se encuentran en el apartado de: Resultados y discusiones.

3.2.4 Shimmer. Se define como la diferencia absoluta promedio entre las amplitudes de periodos consecutivos, dividida entre la amplitud promedio, y se expresa en porcentaje⁴⁹. Generalmente ha sido aplicado para estudiar voces patológicas⁵⁰, sin embargo, también se ha usado para mejorar sistemas de verificación y reconocimiento de hablantes⁵¹.

$$Shimmer (relative) = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A_i - A_{i+1}|}{\frac{1}{N} \sum_{i=1}^N A_i}, \quad (3)$$

donde A_i son los datos de la amplitud pico a pico extraídos y N es el número de periodos extraídos de la frecuencia fundamental.

Se usó una rutina de programación sobre el software Praat para calcular de forma automática el shimmer (relativo) de cada una de las frases pronunciadas por los hablantes. Los resultados de este experimento se encuentran en el apartado de: Resultados y discusiones.

3.2.5 Mean Harmonic to Noise Ratio (HNR). Se calcula como la relación entre la energía de la parte armónica y la energía del resto de la señal en decibeles (dB). Aunque la HNR ha sido generalmente aplicada para analizar voces patológicas, también ha sido usada en algunas aplicaciones relacionadas con la música⁵².

⁴⁹ Ibíd., p. 780.

⁵⁰ ALÍAS, F; SOCORÓ, J y SEVILLANO, X. Op. cit. p. 350.

⁵¹ FARRÚS, M; HERNANDEZ, J y EJARQUE, P. Op. cit., p. 780.

⁵² ALÍAS, F; SOCORÓ, J y SEVILLANO, X. Op. cit. p. 350.

Se usó una rutina del software Praat para calcular de forma automática la HNR de cada una de las frases pronunciadas por los hablantes. El algoritmo usado por Praat para su cálculo realiza una detección de periodicidad acústica basado en un método preciso de autocorrelación⁵³. Los resultados de este experimento se encuentran en el apartado de: Resultados y discusiones.

3.2.6 Envolvente espectral por LPC (Linear Prediction Coefficients).

Representa un filtro todo polo que captura la envolvente espectral de la señal de voz (formantes o resonancias espectrales presentes en el tracto vocal) y ha sido ampliamente usado en aplicaciones de reconocimiento y codificación de la voz⁵⁴. La idea principal de la representación de LPC es considerar la señal de voz $s[n]$ como una combinación lineal de p muestras anteriores de la señal de voz⁵⁵:

$$s[n] = \sum_{k=1}^p a_k s[n-k], \quad (4)$$

donde a_k son los coeficientes de predicción lineal que minimizan la suma de las diferencias cuadráticas entre las actuales muestras de voz y las predichas linealmente⁵⁶. Se usa el algoritmo Levison Durbin para calcular los coeficientes de autocorrelación y posteriormente calcular los coeficientes de predicción⁵⁷.

Los coeficientes de LPC se hallaron con el script `lpc*` de Matlab, el cual pide como entrada la señal de audio y el orden del filtro y da como salida los coeficientes LPC. Así, las señales de audios de entrada fueron las frases completas pronunciadas por 4 de los 10 hablantes, y el orden p fue 12, por lo cual se obtuvieron 13 coeficientes.

⁵³ BOERSMA, P. Accurate short term analysis of the fundamental frequency and the harmonics to noise ratio of a sampled sound. IFA Proceedings, 1993.

⁵⁴ ALÍAS, F; SOCORÓ, J y SEVILLANO, X. Op .cit., p. 350.

⁵⁵ SCHWARZ, D; RODET, X. Spectral Envelope Estimation and Representation for Sound Analysis Synthesis. IRCAM, 1998.

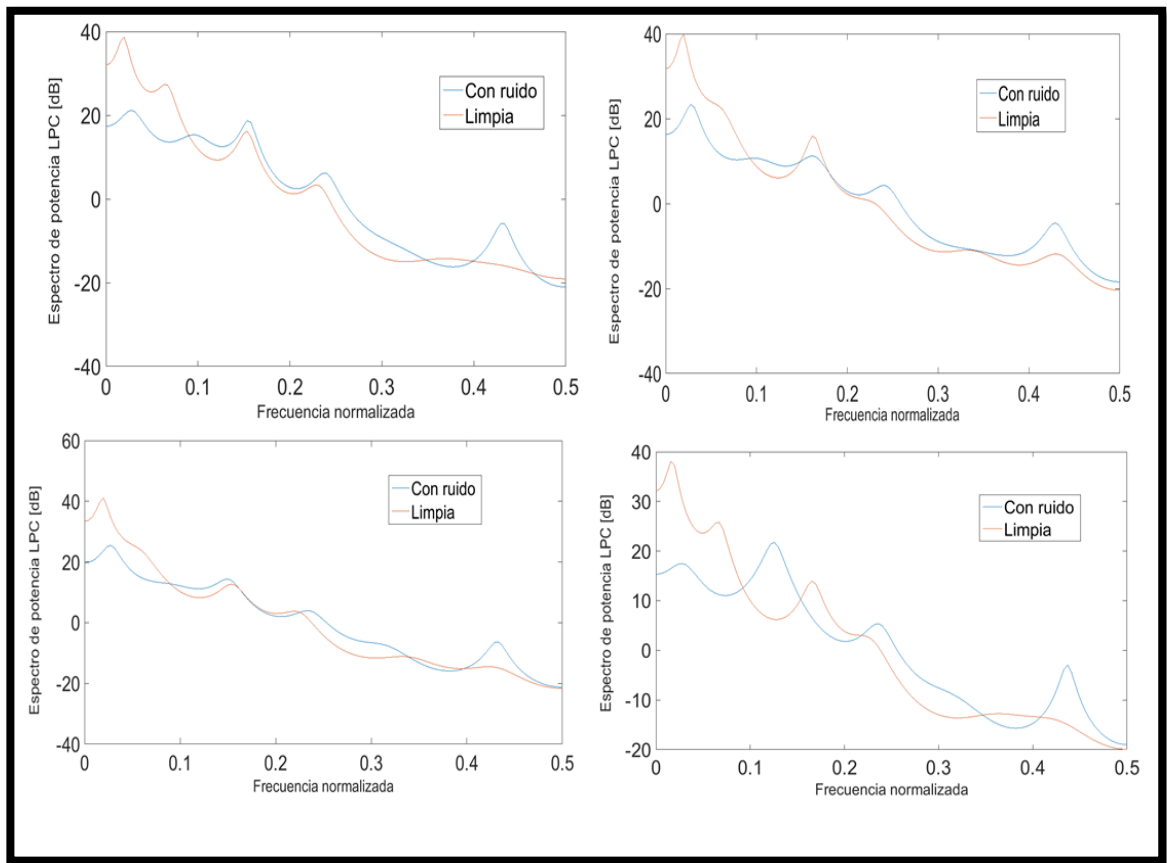
⁵⁶ HAY MAR, S, et al. Comparison of Three Auditory Frequency Scales in Feature Extraction on Myanmar Digits Recognition. 2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE), Kuta, 2018, p. 152.

⁵⁷ Ibid., p. 152.

* El script puede ser descargado desde: <https://www.mathworks.com/help/signal/ref/lpc.html>.

Con la función `lpcar2pf`, la cual pertenece a la toolbox Voicebox** de Matlab, se estimó y graficó la envolvente espectral, con un tamaño de 512 puntos. En la Figura 14 se observa la envolvente espectral de LPC de diferentes frases pronunciadas por los cuatro hablantes de la base de datos. El espectro de color anaranjado corresponde al de la señal de voz limpia (EMA), y el espectro de color azul corresponde a la señal de audio con ruido MRI. Este procedimiento se hizo para todas las frases pronunciadas por los hablantes.

Figura 14. Envoltentes espectrales por LPC de 4 diferentes frases de la base de datos USCM-TIMIT. La envolvente naranja corresponde a la señal de voz limpia y la azul a la señal de voz con ruido MRI.



En la Figura 14 se observa que las diferencias dominantes entre los espectros de las señales de voz limpias y con ruido se presentan en la posición y amplitud de los formantes (picos de los espectros). Estas diferencias son más significativas para el

** La toolbox puede ser descargada desde: <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>.

primer formante, o para el rango de frecuencias comprendidas entre el 0 y 10% de la frecuencia de muestreo.

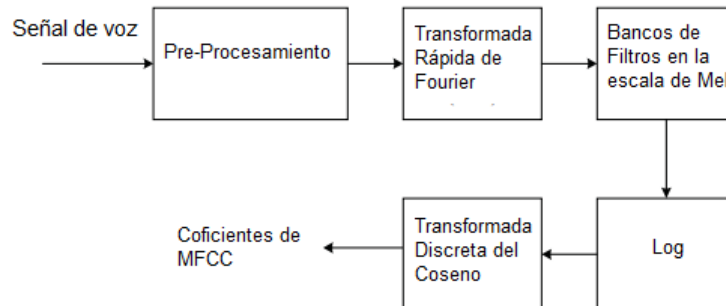
3.2.7 Mel Frequency Cepstrales Coefficients (MFCC) y envolvente espectral.

La escala de Mel es una escala de frecuencias diseñadas para imitar la percepción del oído humano, el cual es más sensible a las señales de audio de baja frecuencia que a las de alta frecuencia. Por lo tanto, para obtener la escala de Mel se diseñan unos filtros triangulares pasabandas que enfatiza las bajas frecuencias sobre las altas frecuencias⁵⁸. La ecuación (5) muestra la conversión de frecuencias de una escala lineal a la escala de Mel⁵⁹.

$$M(f) = 1125 \ln \left(1 + \frac{f}{700} \right) \quad (5)$$

El proceso para la extracción de los MFCC de una señal de voz dada se muestra en la figura 15⁶⁰ :

Figura 15. Diagrama de bloques para la extracción de los MFCC.



Fuente: Propia.

- 1) Se calcula la densidad espectral de potencia
- 2) Se aplica en el dominio de la frecuencia un banco de R filtros en la escala de Mel al espectro de potencia y se obtiene la energía proveniente de cada filtro.

⁵⁸ HAY MAR, S, *et al.* Op. cit., p 154.

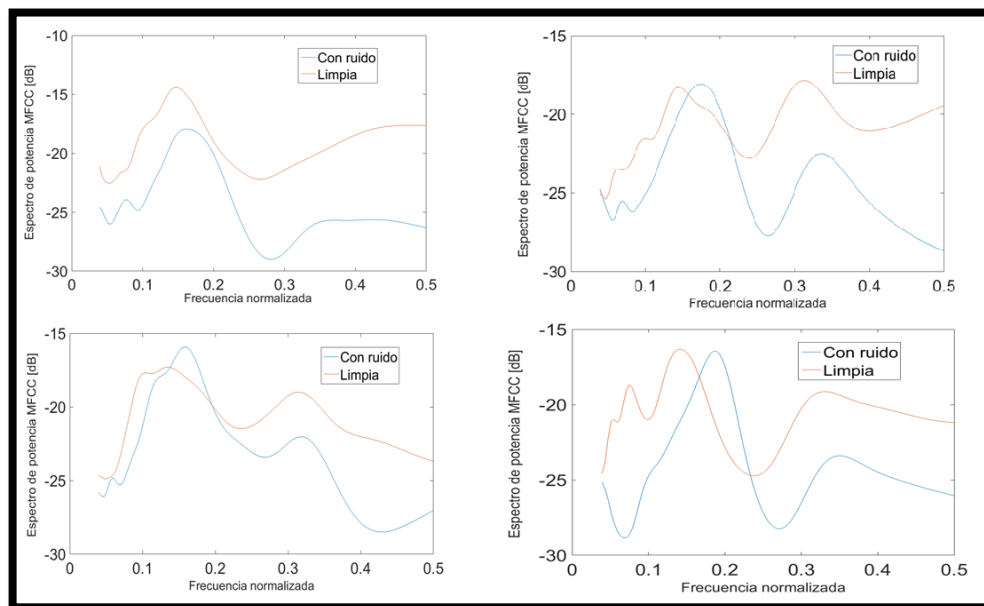
⁵⁹ HUANG, X; ACERO, A y HON, H. Spoken Language Processing: A guide to theory, algorithm, and system development, Prentice Hall, p. 423.

⁶⁰ HAY MAR, S, *et al.* Op. cit., p. 154.

- 3) Se toma el logaritmo de las energías del banco de filtro para formar un vector de longitud R.
- 4) Se aplica una transformación discreta del coseno al vector anterior con el fin de decorrelacionar las energías ya que los filtros se traslapan entre sí⁶¹.

Los coeficientes se hallaron por medio de la función mfcc de Matlab, cuyo autor es Kamil Wojcicki*. Posteriormente, con la función lpcar2pf mencionada anteriormente se estimó y graficó la envolvente espectral, con un tamaño de 512 puntos. En la Figura 16 se observa la envolvente espectral de MFCC de diferentes frases pronunciadas por cuatro hablantes de la base de datos. El espectro de color anaranjado corresponde al de la señal de voz limpia (EMA), y el espectro de color azul corresponde a la señal de audio con ruido MRI. Este procedimiento se hizo para todas las frases pronunciadas por los hablantes.

Figura 16. Envoltentes espectrales por MFCC de 4 diferentes frases de la base de datos USCM-TIMIT. La envolvente naranja corresponde a la señal de voz limpia y la azul a la señal de voz con ruido MRI.



⁶¹ PATINO SAUCEDO, A; SEPULVEDA SEPULVEDA, A y GÓMEZ CAJAS, D. Phoneme recognition system using articulatory-type information. TECCIENCIA. 2015, vol. 10, p. 12.

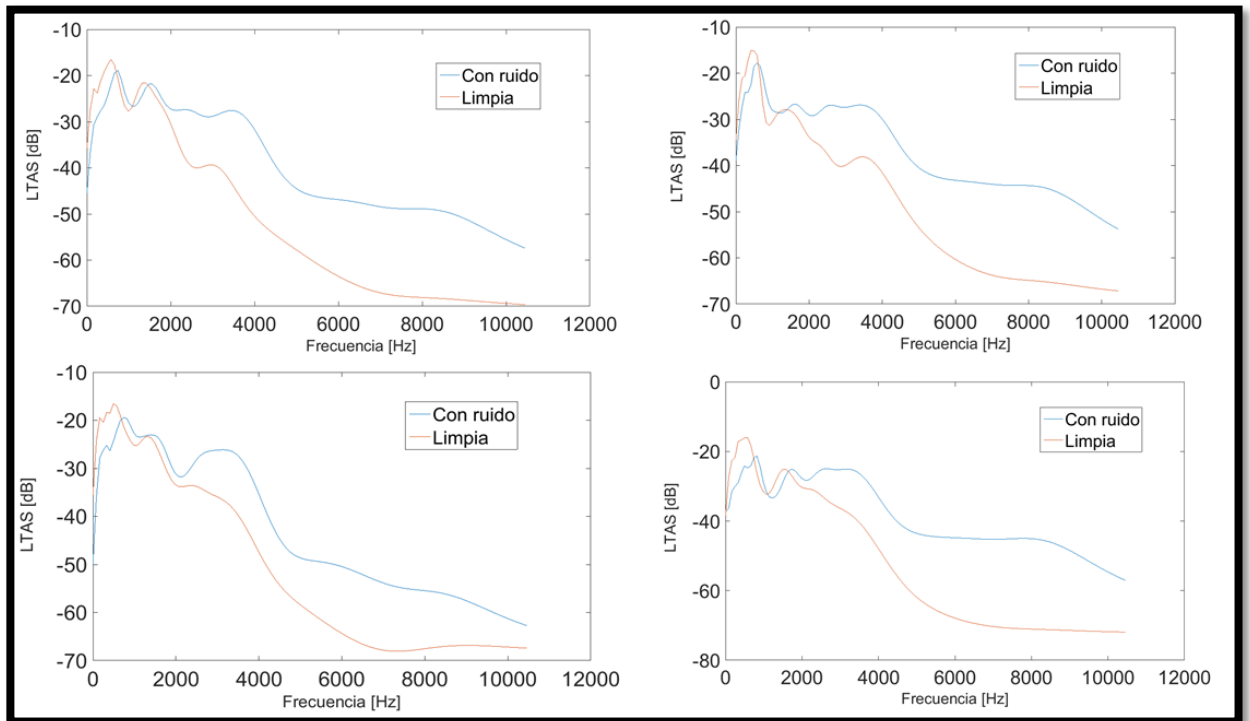
* El script puede ser descargado desde : <https://la.mathworks.com/matlabcentral/fileexchange/32849-htk-mfcc-matlab>

En la Figura 16 se observa que para el espectro de potencia por MFCC de las señales de voz limpias, la mayor parte de la potencia se concentra entre el rango de frecuencia comprendido entre el 5% y 20% de la frecuencia de muestreo. Y para los espectro de potencia por MFCC de las señales de voz con ruido la mayor parte de la potencia se concentra entre el rango de frecuencia comprendido entre el 10% y 30% de la frecuencia de muestreo.

3.2.8 Long Term Average Spectrum (LTAS). Se halló por medio de la función *ltas* que hace parte de la Toolbox IoSR de Matlab*. Los parámetros de entrada para esta función son la señal de audio (que puede ser un vector, una matriz) y la frecuencia de muestreo. El espectro es calculado a partir de la densidad espectral de potencia promedio (DEPP) obtenida de una serie de *Fast Fourier Transforms* (FFT) superpuestas. La longitud de las FFT es de 512 y el tamaño de salto es de 256. Asimismo, los segmentos de la señal de audio de entrada son enventanados por medio de una ventana Hanning. Por último, la DEPP es suavizada a una resolución de 1/3 de octava. En la Figura 17 se observan los LTAS de diferentes frases pronunciadas por cuatro hablantes de la base de datos. El espectro de color anaranjado corresponde al de la señal de voz limpia (EMA) y el espectro de color azul corresponde a la señal de audio con ruido MRI. Este procedimiento se hizo para todas las frases pronunciadas por los hablantes.

* La toolbox puede ser descargada desde: <https://github.com/IoSR-Surrey/MatlabToolbox>

Figura 17. LTAS de 4 diferentes frases de la base de datos USCM-TIMIT. La envolvente naranja corresponde a la señal de voz limpia y la azul a la señal de voz con ruido MRI.



En la Figura 17 se observa que las diferencias entre las envolventes de las señales limpias y con ruido se acentúan después de los 2000 Hz.

Las anteriores figuras correspondientes a las tres representaciones de la señal de voz usadas en el presente trabajo muestran que hay diferencias entre las envolventes espectrales de entre las señales limpias y con ruido. Por ejemplo, las envolventes espectrales no están en fase, una se encuentra desplazada con respecto a la otra. Una posible causa de esta diferencia, además del fuerte ruido acústico, puede ser debido a la misma construcción de los escáneres de MRI; puesto que, los sujetos son ubicados en posición supina, sin excepción, acostados sobre su espalda, y se asume que los efectos gravitacionales son despreciables. No obstante, varios estudios de MRI han observado un desplazamiento hacia atrás

de la lengua y otros efectos posturales que podrían afectar la voz ^{62 63}. Otra posible causa se puede deber a que para adquirir la imágenes o la señal de voz es necesario amplios tiempos de adquisición y según estudios de MRI, el sujeto dentro del escáner a menudo tiene que mantener la articulación artificialmente por 30 s o más⁶⁴. De igual manera hay que tener en cuenta que las habitaciones reverberadas degradan el rendimiento de las técnicas de cancelación de ruido en ambientes de adquisición de imágenes de resonancia magnética funcionales⁶⁵.

3.3 MEDIDAS DE COMPARACIÓN DE SEÑALES DE VOZ

En el procesamiento de voz una medida de distancia $d(c, \hat{c})$, donde c y $\hat{c}$ son dos representaciones acústicas particulares, debe satisfacer por los menos las siguientes propiedades⁶⁶ :

$$\begin{array}{ll} \text{Simetría} & d(c, \hat{c}) = d(\hat{c}, c) \\ \text{definición positiva} & \begin{cases} d(c, \hat{c}) > 0 \forall c \neq \hat{c} \\ d(c, \hat{c}) = 0 \text{ si } c = \hat{c} \end{cases} \end{array} \quad (6)$$

De lo anterior se concluye que la distancia de c a $\hat{c}$ es igual a la distancia de $\hat{c}$ a c , y que la distancia entre c y $\hat{c}$ es positiva, excepto cuando $c = \hat{c}$, en este caso la distancia es cero. Asimismo, el requisito de simetría asegura que la medida de distancia entre fragmentos de sonidos no tenga que distinguir entre cual es el fragmento de referencia y cual el de prueba.

Por otro lado, las distancias L^p son comúnmente usadas en ingeniería y están definidas por la siguiente expresión⁶⁷:

⁶² ENGWALL, O. A Revisit to the Application of MRI to the Analysis of Speech Production-Testing our Assumptions. Sydney, 2003.

⁶³ STONE, M. Comparison of speech production in upright and supine position. Acoustical Society of America. 2007, p. 532.

⁶⁴ RAMANARAYANAN, V, et al. Op. cit., p. 225.

⁶⁵ BAO, H; PANAH, I y BRIGGS, R. Blind dereverberation for fMRI noise based on SIMO linear prediction method. IEEE. 2007, p. 2832.

⁶⁶ SEPULVEDA, Alexander. Op. cit., p. 13.

⁶⁷ E. K. Introductory Functional Analysis with Applications. John Wiley & Sons, 1989.

$$d(\mathbf{c}, \hat{\mathbf{c}}) = \left(\sum_k |c_k - \hat{c}_k|^p \right)^{\frac{1}{p}}, \quad (7)$$

donde c_k y $\hat{c}_k$ son las entradas de los vectores $\mathbf{c}, \hat{\mathbf{c}}$, respectivamente. Las medidas L^p , para $p \geq 1$, son verdaderas medidas de distancia en el sentido que satisfacen la inecuación triangular $d(\mathbf{c}, \hat{\mathbf{c}}) \leq d(\mathbf{c}, \mathbf{c}') + d(\mathbf{c}', \hat{\mathbf{c}})$ ⁶⁸, además de cumplir con las definiciones de simetría y positiva. A continuación, se definirán las medidas de distancia espectrales utilizadas en el presente trabajo de investigación:

3.3.1 Medidas log espectrales. Considere dos modelos espectrales $P_{(w)}$ y $\hat{P}_{(w)}$, el error o la diferencia entre estos modelos en magnitud logarítmica (log) versus una escala frecuencial está definida por⁶⁹:

$$E(w) = \log(P(w)) - \log(\hat{P}(w)) \quad (8)$$

Un conjunto de opciones lógicas para una medida de distancia entre los modelos espectrales es el conjunto de normas L^p , o medidas d , definidas en la ecuación (7), entonces se obtiene:

$$d(P_{(w)}, \hat{P}_{(w)}) = \left(\frac{1}{N} \sum_{k=1}^N |E(w)_k|^p \right)^{\frac{1}{p}}, \quad (9)$$

donde N es la longitud de cada uno de los modelos espectrales. Para $p = 1$, se define la medida log espectral absoluta promedio⁷⁰. Para $p = 2$, se define la medida rms log espectral. Para $p = \infty$, se obtiene la diferencia pico log espectral.

3.3.2 Coeficiente de correlación espectral de Pearson (P). Este coeficiente, es la relación lineal entre dos variables y puede ser calculado a partir de la variación total de un determinado conjunto de datos y puede ser escrito como⁷¹:

⁶⁸ KOLOMOGOROV, A; FOMIN, S. Introductory Real Analysis. Dover Publications, 1970.

⁶⁹ A, Gray; M, John. Distance Measures for Speech Processing. IEEE. 1976, vol. 4, nº 5, pp. 380.

⁷⁰ Ibid., p. 381.

⁷¹ SPIEGEL, M. Theory and problems of Statistics in SI units. Schaum's Outline Series, 1972.

$$C_{correlation} = \frac{\sum_{k=1}^N (f_{kT} - \hat{f}_T)(f_{kR} - \hat{f}_R)}{\sqrt{\sum_{k=1}^N (f_{kT} - \hat{f}_T)^2 \sum_{k=1}^N (f_{kR} - \hat{f}_R)^2}}, \quad (10)$$

donde f_{kT} y f_{kR} son las bandas espectrales de las muestras de prueba y de referencia respectivamente; y $\hat{f}_T$ y $\hat{f}_R$ son sus respectivas medias. El valor del coeficiente varía entre -1 y +1; +/- 1 indica una perfecta correlación y 0 quiere decir que no hay correlación.

Aunque la correlación no es una medida de distancia sirve para medir el grado de similitud entre estimaciones y señales.

3.3.3 Log Likelihood Ratio (LLR). El test de log likelihood ratio compara de forma global la envolvente espectral de la señal de voz limpia y mejorada; y está definida como⁷²:

$$d_{LLR}(a_c, a_e) = \log \frac{a_e^T R_c a_e}{a_c^T R_c a_c}, \quad (11)$$

donde a_c es un vector de coeficientes de predicción lineal de la señal de voz limpia, a_e es un vector de coeficientes de predicción lineal de la señal de voz mejorada, y R_c es la matriz de autocorrelación de la señal de voz limpia. Valores muy grandes de d_{LLR} indican una mayor distorsión en los segmentos de voz de la señales de audio. Se halló con la función composite de Matlab cuyo autor es Philipos C. Loizou⁷³.

3.4 MEDIDAS OBJETIVAS DE LA CALIDAD DE LA VOZ:

3.4.1 Medidas compuestas CSIG, CBAK, COVL. En ⁷⁴ desarrollaron tres medidas objetivas compuestas para evaluar la calidad de la voz: **csig**, **cbak** y **covl**. Estas medidas están basadas en la metodología de prueba subjetiva para evaluar los

⁷² MONTAZERI,V; PATHAK,N y PANAHI,I. Op. cit., p. 65.

⁷³ LOIZOU, P. Speech Enhancement: theory and Practice, Boca Raton, 2007.

⁷⁴ HU, Y; LOIZOU, P. Evaluation of Objective Quality Measures for Speech Enhancement. IEEE Transactions on audio, speech, and language processing. Enero, 2008, vol. 16, nº 1, pp. 235.

sistemas de comunicación vocal que utilizan un algoritmo de cancelación de ruido, ITU-T P8.35*. Dicha metodología evalúa la calidad de la voz filtrada de acuerdo a tres dimensiones: calidad subjetiva de la señal de voz en sí (SIG), el ruido de fondo o ambiente (BAK) y la calidad general (OVL). Estas tres medidas cuentan con una escala de 5 puntos que se pueden observar en la Tabla 2.

Tabla 2. Descripción de la escala de las tres dimensiones (SIG, BAK, OVL) usada en la metodología ITU-T P8.35.

Escala	SIG	BAK	OVL
5	Sin distorsión	Imperceptible	excelente
4	Ligeramente distorsionada	Ligeramente perceptible	buena
3	Algo distorsionada	perceptible aunque no molesto	Regular
2	Bastante distorsionada	Algo molesto	Mediocre
1	Muy distorsionada	Muy molesto	Mala

Fuente: HU, Y; LOIZOU.

Las tres medidas objetivas compuestas fueron obtenidas de la combinación de tres medidas objetivas básicas: segmental SNR (segSNR), weighted-slope spectral distance (WSS), PESQ, y log-likelihood ratio (LLR). Las medidas compuestas son necesarias ya que no se puede esperar que las medidas objetivas convencionales (por ejemplo, LLR) se correlacionen altamente con las distorsiones debido a las tres dimensiones, SIG, BAK, OVL⁷⁵. Además, se pueden derivar al utilizar análisis de regresión lineal múltiples o por la aplicación de técnicas no lineales como por ejemplo, multivariable adaptive regression splines (MARS)⁷⁶. La tarea para formar una buena medida compuesta por la combinación lineal de medidas objetivas no es fácil. Idealmente, se esperaría combinar medidas objetivas que se correlacionen altamente con las escalas de los test subjetivos y que a la vez capturen las

* La norma ITU-T-P835 se puede descargar desde la siguiente dirección web: <https://www.itu.int/rec/T-REC-P.835>

⁷⁵ HU, Y; LOIZOU, P. Op. cit., p. 235.

⁷⁶ Ibid., p. 235.

diferentes características de la distorsión presente en la señal de voz mejorada. No existe un método sencillo para seleccionar el mejor conjunto de medidas objetivas para usar en las medidas compuestas, aparte de probar diferentes combinaciones y evaluar la correlación resultante⁷⁷.

3.4.2 Perceptual Evaluation of Speech Quality (PESQ). PESQ compara una señal inicial $X(t)$ con una señal degradada $Y(t)$. La salida de PESQ es una predicción de la calidad percibida por los sujetos en una prueba de escucha subjetiva y que sería atribuida a $Y(t)$. Así, para evaluar la calidad de la voz mejorada, el algoritmo alinea la voz limpia y la voz mejorada en el dominio del tiempo y luego pasa ambas a través de un modelo perceptual y compara las salidas respectivas entre sí. Finalmente, las salidas del modelo perceptivo se filtran mediante un sistema que representan el modelo cognitivo. Para más detalles del algoritmo se puede revisar⁷⁸. La puntuación de PESQ varía entre 0.5 (pobre calidad) a 4.5 (excelente calidad)⁷⁹.

La puntuación PESQ se calcula como una combinación lineal del valor de perturbación promedio D_{ind} y los valores de perturbación asimétrica promedio A_{ind} de la siguiente manera⁸⁰:

$$PESQ = a_0 + a_1 D_{ind} + a_2 A_{ind} \quad (12)$$

donde $a_0 = 4.5$, $a_1 = -0.1$, y $a_2 = -0.0309$. Los parámetros a_0 , a_1 y a_2 fueron optimizados por algoritmos de redes y no por algoritmos de supresión de ruido. Como en ⁸¹ no esperaban que la medida PESQ se correlacionara altamente con las tres dimensiones de calidad (SIG, BAK y OVL), optimizaron la medida PESQ para cada una, eligiendo un conjunto diferente de parámetros (a_0 , a_1 , a_2) para cada dimensión de calidad (Tabla 3); un análisis múltiple de regresión lineal fue usado para determinar los 3 parámetros. Los valores de D_{ind} y A_{ind} en la ecuación (12)

⁷⁷ Ibid., p. 235.

⁷⁸ ITU_T Recommendation, P.862. Op. cit.

⁷⁹ MONTAZERI, V; PATHAK, N y PANAHI, I. Op. cit., p. 64.

⁸⁰ HU, Y; LOIZOU, P. Op. cit., p. 235.

⁸¹ Ibid., p. 236.

fueron tratados como variables independientes en el análisis de regresión. Las actuales escalas subjetivas presentadas en la Tabla 2 también fueron usadas en el análisis de regresión. Este análisis arrojó tres diferentes medidas PESQ modificadas adecuadas para predecir la distorsión de la señal (SIG), la distorsión del ruido (BAK), y la calidad general de la voz (OVL).

Tabla 3. Conjunto de coeficientes a0, a1, a2 de PESQ para cada dimensión de calidad.

Dimensión	a0	a1	a2
SIG	4.959	-0.191	0.006
BAK	5.336	-0.082	-0.058
OVL	4.788	-0.152	-0.016

Fuente: HU, Y; LOIZOU.

4. RESULTADOS Y DISCUSIÓN

Se trabajó con los archivos de audio EMA y MRI de 4 de los 5 hablantes que tienen disponibles archivos EMA. El presente capítulo se divide en tres secciones:

- La evaluación de la calidad de las señales de voz obtenidas con el sistema de adquisición de MRI, de acuerdo a metodologías de pruebas subjetivas y objetivas que evalúan los sistemas de comunicación vocal que utilizan un algoritmo de cancelación de ruido.
- La comparación de la envolvente espectral de la señal de voz limpia (EMA) y de la envolvente espectral de voz con ruido (MRI), por medio de medidas de distancia y estadísticas.
- Análisis de la influencia de los sistema de MRI sobre medidas acústicas: Fo, jitter, shimmer, HNR y los primeros cuatro formantes.

Evaluación de la calidad de la voz: Se evaluó el rendimiento de la metodología de reducción de ruido del sistema de adquisición de MRI usado para obtener las señales de voz de la base de datos USC-TIMIT. Para ello se obtuvieron medidas de calidad de la voz por medio de las siguientes medidas objetivas: csig, cbak, covl, LLR y PESQ (Capítulo 2, sección 5); fueron calculadas por medio de la rutina escrita en Matlab por Philipos C. Loizou y Yi Hu Brayan⁸². En la Tabla 4 se observan el promedio y desviación estándar de los resultados de cada medida para los 4 hablantes evaluados.

Tabla 4. Resultados de las medidas para la evaluación objetiva de la calidad de la voz.

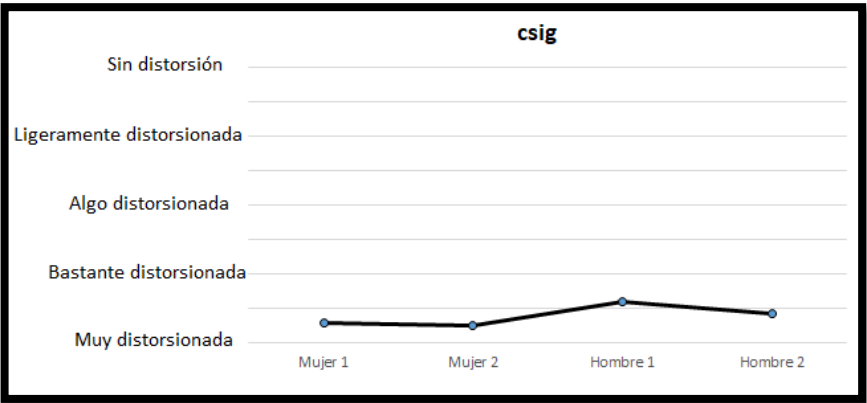
Hablante	csig [1-5]	cbak [1-5]	covl [1-5]	LLR [0-2]	PESQ [0.5-4.5]
Mujer 1	1.28 $\sigma=0.27$	1.47 $\sigma=0.22$	1.32 $\sigma=0.26$	0.99 $\sigma=0.41$	1.92 $\sigma=0.37$
Mujer 2	1.25 $\sigma=0.22$	1.42 $\sigma=0.21$	1.34 $\sigma=0.23$	0.73 $\sigma=0.44$	1.7 $\sigma=0.29$
Hombre 1	1.59 $\sigma=0.46$	1.51 $\sigma=0.28$	1.75 $\sigma=0.46$	0.94 $\sigma=0.3$	1.77 $\sigma=0.51$
Hombre 2	1.42 $\sigma=0.31$	1.6 $\sigma=0.2$	1.72 $\sigma=0.32$	0.73 $\sigma=0.35$	1.92 $\sigma=0.39$

⁸² LOIZOU, P. Op. cit., p. 229-238.

A continuación se presenta la relación de los resultados presentados en la Tabla 6 y la puntuación subjetiva de la calidad presentada en la Tabla 4 (capítulo II, sección 5)

Csig:

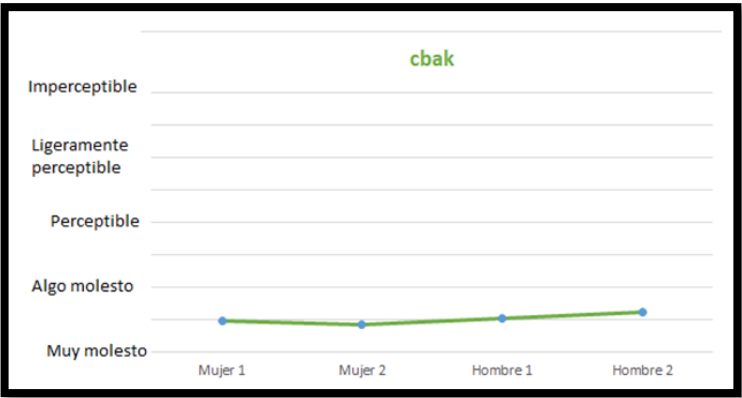
Figura 18. Relación entre los resultados de csig y la escala SIG.



Según la Figura 18, las señales de voz adquiridas con el sistema de MRI, de los cuatro hablantes analizados de la base de datos USCM-TIMIT, están entre muy distorsionadas y bastante distorsionadas.

Cbak:

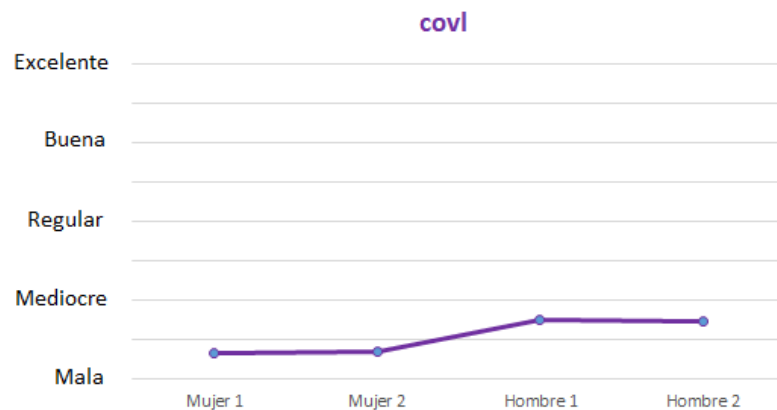
Figura 19. Relación entre los resultados de csig y la escala SIG.



Según la Figura 19, el ruido debido a los dispositivos usados en los sistemas de adquisición de MRI, presente en las señales de voz de los cuatros hablantes analizados, está entre muy molesto y algo molesto.

Covl:

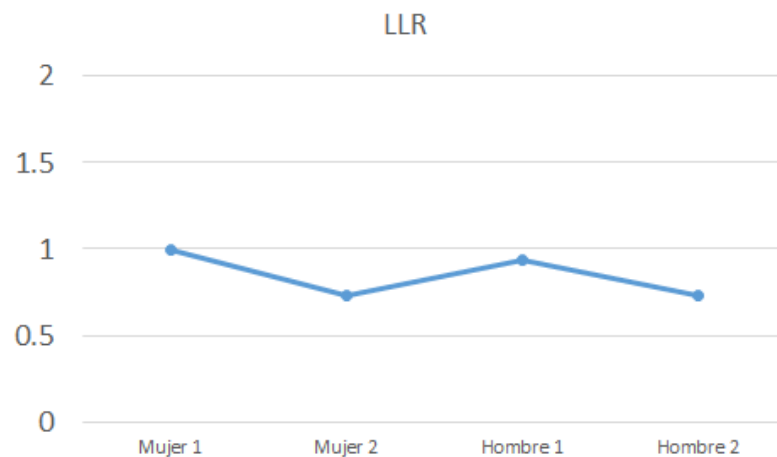
Figura 20. Relación entre los resultados de covl y la escala OVL



Según la Figura 20, la calidad en general de las señales de voz adquiridas con el sistema de MRI, de los cuatro hablantes analizados, está entre mala y mediocre.

LLR:

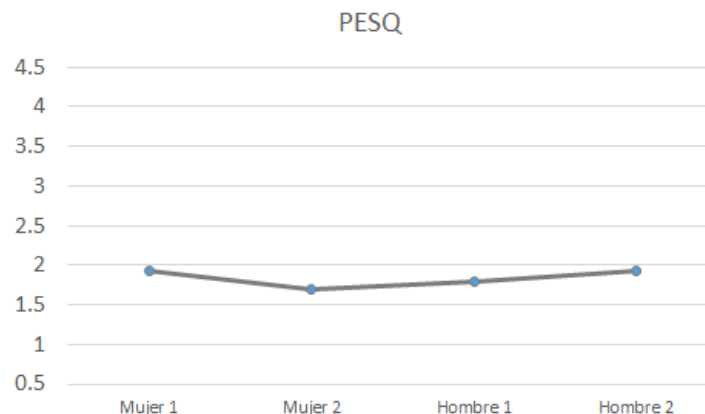
Figura 21. Resultados de la medida LLR



En la Figura 21 se observa que el valor promedio de LLR, entre las señales de voz adquiridas con el sistema de MRI, y las señales de voz adquiridas con EMA, de los 4 hablantes, es cercano a 1, es decir, la mitad del máximo valor de distorsión, lo cual indica que es muy probable que aproximadamente el 50% de variabilidad observada entre los registros de audio de MRI y los de EMA se deba a ruidos parásitos presentes en los segmentos de voz de las señales de audio adquiridas con MRI.

PESQ:

Figura 22. Resultados de la medida PESQ.



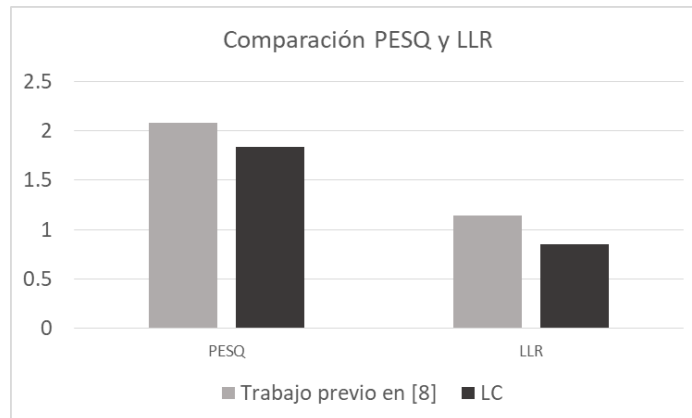
En la Figura 22 se observan que los resultados obtenidos de la medida PESQ, entre las señales de voz adquiridas con el sistema de MRI y las señales de voz adquiridas con EMA, de los 4 hablantes, no alcanza siquiera a ser de 2, es decir, la percepción de la calidad de las señales de voz con ruido de MRI de la base de datos USCM – TIMIT no es medianamente aceptable.

Posteriormente se obtuvo un único promedio PESQ y LLR a partir de los resultados de los cuatros hablantes y se compararon con los resultados obtenidos en⁸³, en la Figura 23 se puede observar la comparación. La barras de color gris corresponde a

⁸³ MONTAZERI, V; PATHAK, N y PANAHI, I. Op. cit., p. 65.

los valores PESQ y LLR obtenidos con uno de los métodos de reducción de⁸⁴, y las barras negras a los del este trabajo de grado.

Figura 23. Comparación entre los valores de PESQ y LLR.



Si se relacionan estos valores de PESQ con la tercera columna de la escala subjetiva de la Tabla 4, se encuentra que la calidad de la voz resultante no alcanza a ser regular, por lo tanto se puede decir que el diseño de algoritmos o técnicas de supresión de ruidos integrados a los sistemas de MRI es todavía un campo que está en proceso de desarrollo.

Envolvente espectral: Para comparar las representaciones de las envolventes espectrales obtenidas a partir de los LPC, MFCC y LTAS, de las señales de voz limpias (EMA) y con ruido, se calcularon las distancias log espectrales ($p = 1$ y $p = 2$). En la Tabla 5 aparecen la media (μ) y desviación estándar (σ) de las dos distancias espectrales usadas sobre las tres formas diferentes de representación espectral:

Hay que aclarar que aquí lo que se busca es comparar las tres diferentes representaciones espectrales y ver cual se ve menos afectada por el ruido de MRI.

⁸⁴ Ibid., p. 69.

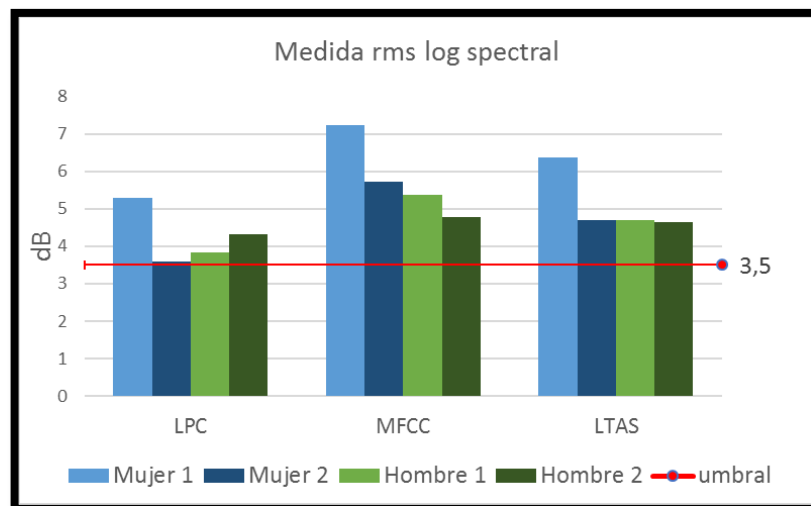
Tabla 5. Resultado de las dos medidas log espectrales (ec 9) para comparar las tres representaciones espectrales.

$p=1 [dB]$								
Envolvente	Mujer 1		Mujer 2		Hombre 1		Hombre 2	
	μ	σ	μ	σ	μ	σ	μ	σ
LPC	4.01	0.46	2.82	0.34	2.92	0.8	3.49	0.64
MFCC	4.77	0.73	4.4	0.82	3.86	0.9	2.9	0.79
LTAS	4.53	0.65	3.73	0.62	3.49	0.68	3.29	0.55
$p=2 (rms) [dB]$								
Envolvente	μ	σ	μ	σ	μ	σ	μ	σ
LPC	5.28	0.58	3.52	0.32	3.83	1.02	4.31	0.7
MFCC	7.24	0.89	5.71	0.83	5.36	0.83	4.78	0.92
LTAS	6.38	0.66	4.71	0.51	4.69	0.71	4.64	0.81

Para la medida log espectral rms por debajo de 0.5 dB entre envolventes espectrales no se pueden percibir cambios en la señal de voz; entre 0.5 y 2 dB los cambios serán apenas perceptibles; y por encima de los 3.5 dB los cambios ya serán perceptibles⁸⁵.

En la Figura 24 se puede observar mejor esta comparación, la línea roja horizontal corresponde al umbral de 3.5 dB:

Figura 24. Comparación entre la medida rms log espectral (ec.9,p=2) de las tres representaciones de la envolvente espectral de cada hablante, con el umbral de distorsión.



⁸⁵ A, Gray; M, John. Op. cit., p. 388.

Adicionalmente, se calculó el coeficiente de correlación de Pearson dado por la ecuación (9). Los resultados se presentan en la Tabla 6:

Tabla 6. Coeficientes de correlación de Pearson (ec 10) de las tres representaciones de la envolvente espectral.

Hablante	LPC	MFCC	LTAS
Mujer1	0.72	0.49	0.73
Mujer 2	0.81	0.52	0.81
Hombre 1	0.88	0.57	0.93
Hombre 2	0.95	0.59	0.94

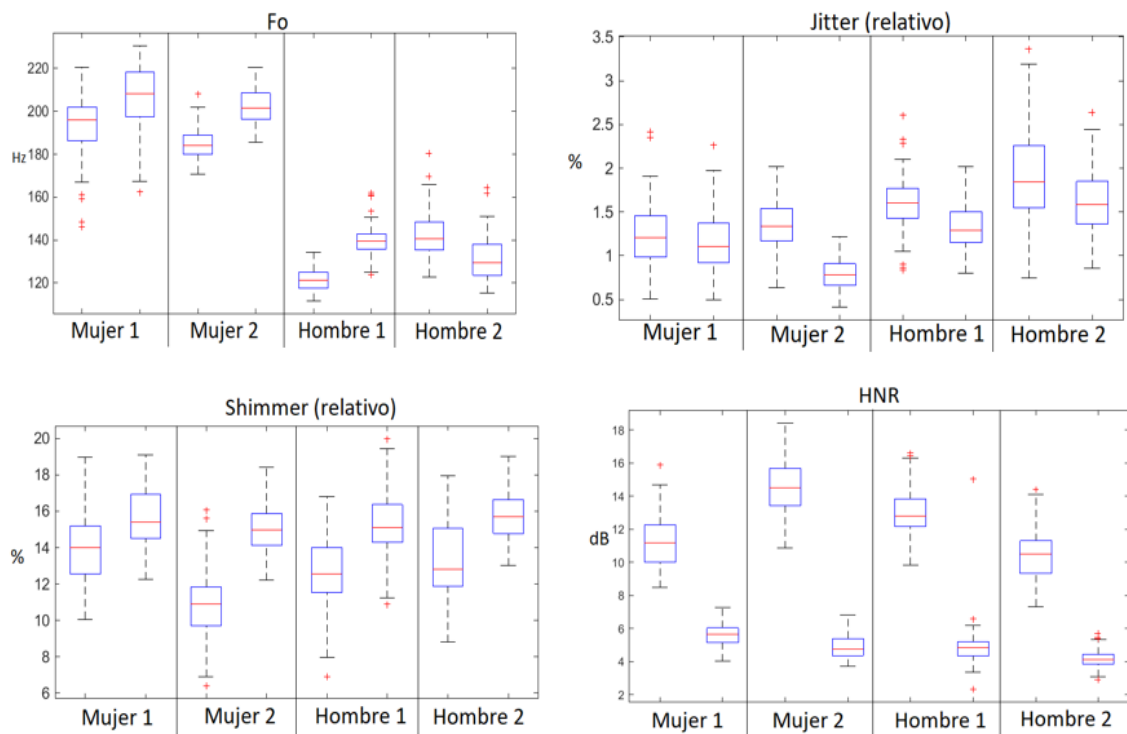
Los resultados presentados en la Tabla 5, Tabla 6 y la Figura 24 muestran que la adquisición de MRI si afecta la percepción auditiva de la señal de voz, puesto que los promedios de la distancia rms log espectral de las representaciones por LPC, MFCC, LTAS, entre las señales de audio limpias y con ruido de MRI están por encima del umbral de los 3.5 dB. No obstante, la diferencia respecto a este umbral no es muy grande: para el espectro de LPC la diferencia en promedio es de 1.2 veces el valor de cambios ya perceptibles; para MFCC es 1.6 veces; y el de LTAS la diferencia en promedio es de 1.5 veces el valor de cambios ya perceptibles.

Entonces, si se desea evitar la influencia del ruido se recomienda el uso de una representación espectral por LPC, porque además de presentar la menor distancia, cuenta con una alta similitud entre espectro de señales limpias y sin ruido, como lo comprueba su correspondiente coeficiente de Pearson (Tabla.6). Por otra parte, aunque la medida de distancia para la representación espectral por MFCC es muy cercana a la de LPC, el promedio del valor del coeficiente de Pearson muestra que esta representación se ve más afectada por el ruido.

Medidas acústicas: Fo, jitter, shimmer, HNR, formantes: A cada frase tanto limpia como con ruido pronunciada por los cuatro hablantes se les halló su Fo, jitter,

shimmer, HNR, y los primeros 4 formantes (F1, F2, F3, F4). Para ello se utilizó una rutina de programación sobre el software Praat (como ya se mencionó en la sesión 3 del capítulo II). La Figura 25 y en La Figura 26 corresponden a los boxplots de los datos de las medidas acústicas. Los boxplot del lado izquierdo son los valores de las medidas provenientes de las señales de voz limpia; y las cajas del lado derecho son los valores de las medidas provenientes de las señales de voz con ruido.

Figura 25. Boxplots de las medidas acústicas Fo, jitter, shimmer y HNR : Las cajas del lado izquierdo pertenece a los datos provenientes de las señales de voz limpia, y los diagramas de caja del lado derecho pertenecen a los datos de provenientes de las señales de voz con ruido.

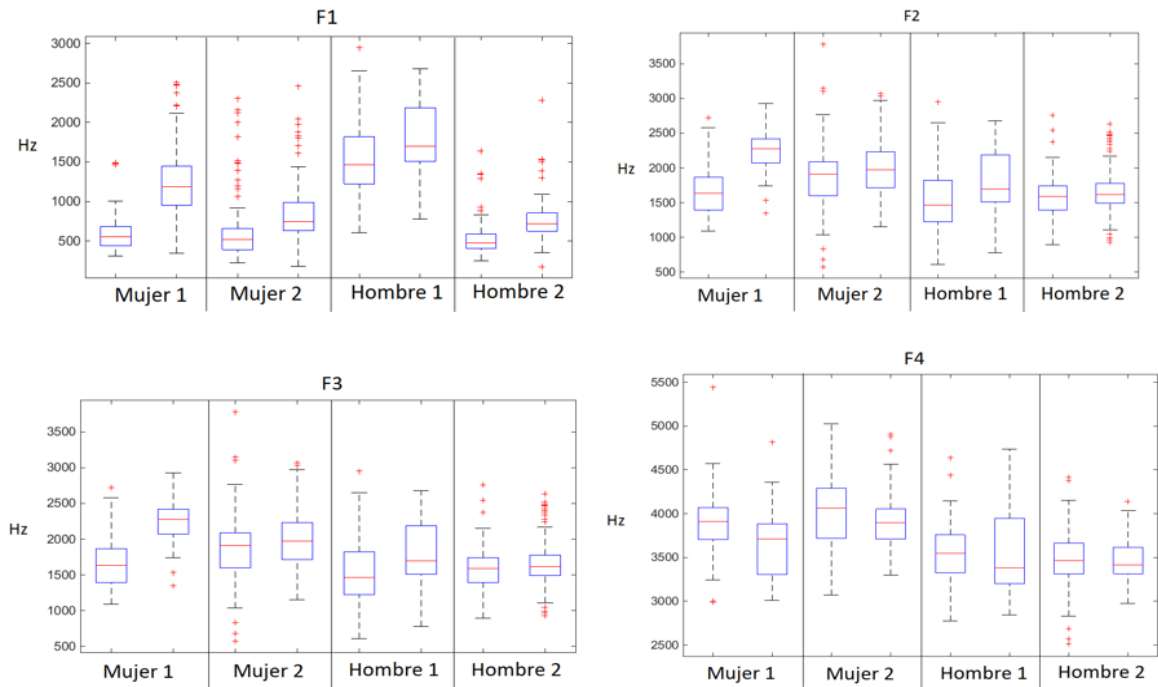


De los boxplots presentados en la Figura 25 se observa que:

- De forma general las medianas de las medidas acústicas provenientes de las señales limpias y con ruido son distintas.

- En la mayoría de los casos la dispersión de las medidas acústicas provenientes de las señales limpias y con ruido difieren pero no considerablemente.
- Para Fo se puede decir que los datos de las muestras de las señales limpias y señales con ruido de la mujer 2 son diferentes entre sí; lo mismo los del hombre 1; pero no podemos asegurar lo mismo de los datos de las muestras de la mujer 1 los del hombre 2.
- Para el jitter se puede decir que los datos de las muestras de las señales limpias y señales con ruido de la mujer 2 son diferentes entre sí; pero no podemos asegurar lo mismo de los de los otros tres hablantes.
- Para el shimmer se puede decir que los datos de las muestras limpias y con ruido de la mujer 2 son diferentes entre sí; lo mismo los del hombre 1; pero no podemos asegurar lo mismo de los datos de la mujer 1 el hombre 2.
- Para el HNR se puede decir que los datos de las muestras limpias y con ruido de los cuatros hablantes son diferentes entre sí.

Figura 26. Boxplot de los cuatro formantes (F1, F2, F3, F4): Las cajas del lado izquierdo pertenece a los datos de los formantes provenientes de las señales de voz limpia, y los diagramas de caja del lado derecho pertenecen a los datos de los formantes provenientes de las señales de voz con ruido.



De los boxplots presentados en la Figura 26 se observa que:

- De forma general las medianas de los formantes provenientes de las señales limpias y con ruido son distintas.
- En la mayoría de los casos la dispersión de los formantes provenientes de las señales limpias y con ruido difieren pero no considerablemente.
- Para el F1 se puede decir que los datos de las muestras de las señales limpias y señales con ruido de la mujer 1, mujer 2, hombre 2 son diferentes; pero no podemos asegurar lo mismo de los datos de las muestras limpias y con ruido del hombre 2.
- Para el F2 se puede decir que los datos de las muestras de las señales limpias y señales con ruido de la mujer 1 son diferentes; pero no podemos asegurar lo mismo de los de los otros tres hablantes.

- Para el F3 se puede decir que los datos de las muestras limpias y con ruido de la mujer 1 son diferentes; pero no podemos asegurar lo mismo de los datos de las muestras de los otros tres hablantes.
- Para el F4 no se puede decir que los datos de las muestras limpias y con ruido de los cuatros hablantes son diferentes.

Pruebas estadísticas para comparación de poblaciones: De la Figura 25 y Figura 26 se observa que para los cuatro hablantes, los datos obtenidos de las medidas acústicas proveniente de las frases con ruido y de los datos obtenidos de las medidas acústicas de las frases limpias, difieren entre ellas con respecto a la mediana, variación y función de distribución. Para algunas medidas acústicas y algunos hablantes estas diferencias son bastantes evidentes y para otros no tanto. Por lo tanto, es necesario el uso de pruebas de hipótesis para validar o rechazar, que efectivamente, las medidas acústicas de los archivos de audio con ruido y de los archivos de audio limpios tienen diferentes medias, varianzas y funciones de distribución:

De la varianza: A continuación, se realiza una prueba de tipo estadístico con el fin de aceptar o descartar la hipótesis nula (H_0) que dos poblaciones, medidas acústicas con ruido y medidas acústicas sin ruido, comparten la misma varianza. La prueba de hipótesis F de dos muestras es usada para probar si las varianzas de dos poblaciones son iguales. El estadístico está dado por ⁸⁶ :

$$F = \frac{s_1^2}{s_2^2}, \quad (13)$$

⁸⁶ PRESS, W, et al. Numerical Recipes in C: The art of scientific computing, Cambridge University Press, 1992, p. 550.

donde s_1 y s_2 son las desviaciones estándar muestrales. La prueba estadística es una relación entre las dos varianzas muestrales. Cuanto más se desvíe esta relación de 1, más probable es que la hipótesis nula sea rechazada. Sometido a la hipótesis nula, el estadístico F tiene una distribución- F con grados de libertad en el numerador igual a N_1-1 , y grados de libertad en el denominador igual a N_2-1 , donde N_1 y N_2 son los tamaños muestrales de los dos conjuntos de datos.

La prueba de hipótesis- F de dos muestras se halló con la función *vartest2* de Matlab. La decisión de la prueba es 1 si H_0 es descartada con un nivel de significancia del 5% y 0 en caso contrario, es decir, que las muestras comparadas tienen distintas varianzas.

El valor de H_0 fue de 1 para todas las medias acústicas comparadas en el presente trabajo de grado, por lo tanto se descarta que tengan la misma varianza.

De las medias: A continuación, se realiza una prueba de tipo estadístico con el fin de comparar si dos poblaciones, medidas acústicas con ruido y medidas acústicas sin ruido, comparten la misma media. Se usó la prueba de hipótesis t de dos muestras. Sean X_1, X_2 dos muestras independientemente distribuidas, con tamaños n_1 y n_2 , medias μ_1 y μ_2 , varianzas σ_1^2 y σ_2^2 , respectivamente, donde $n_i (i = 1, 2)$ son conocidas y μ_i, σ_i son desconocidas. La hipótesis nula H_0 y la hipótesis alternativa H_1 son⁸⁷:

$$H_0: \mu_1 = \mu_2; H_1: \mu_1 \neq \mu_2, \quad (14)$$

⁸⁷ Ibíd., p. 552.

y el estadístico usado en la prueba- t de dos muestras está dado por⁸⁸:

$$t_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}, \quad (15)$$

$$v = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{(s_1^2)^2/(n_1 - 1) + (s_2^2)^2/(n_2 - 1)}, \quad (16)$$

donde $\bar{X}_1, \bar{X}_2$ son las medias de X_1, X_2 ; s_1^2, s_2^2 son las varianzas muestrales correspondiente a X_1, X_2 ; v es el grado de libertad de la distribución- t . Dado un nivel de significancia α , la hipótesis nula H_0 será descartada si $t_0 > t_{\alpha, v}$, donde $t_{\alpha, v}$, es el punto crítico α superior de la distribución- t con v grados de libertad.

La prueba de hipótesis- t de dos muestras se halló con la función `ttest2` de Matlab. La decisión de la prueba es 1 si H_0 es descartada con un nivel de significancia del 5% y 0 en caso contrario, es decir, que las muestras comparadas tienen distintas medias

El valor de H_0 fue de 1 para todas las medias acústicas comparadas en el presente trabajo de grado, por lo tanto se descarta que tengan la misma media.

Función de distribución: A continuación, se realiza una prueba de tipo estadístico con el fin de comparar si dos poblaciones, medidas acústicas con ruido y medidas acústicas sin ruido, provienen de la misma función de densidad de probabilidad.

⁸⁸ *Ibíd.*, p. 553.

Se usó la prueba de hipótesis de dos muestras de Kolmogorov-Smirnov, la cual compara dos funciones de distribución empírica (fde)⁸⁹. Para esto se plantea el siguiente estadístico:

$$D_{n,m} = \max |F_n(x) - G_m(x)|, \quad (17)$$

donde $F_n(x)$ y $G_m(x)$ son las fde del primer y segundo vector de datos respectivamente. La hipótesis nula (H_0) de esta prueba es que los dos vectores de datos provienen de la misma distribución continua, y la hipótesis alternativa es que son de diferentes distribuciones continuas.

La hipótesis nula es rechazada en el nivel de significancia (α) si cumple que⁹⁰:

$$D_{n,m} > c(\alpha) \sqrt{\frac{n+m}{nm}}, \quad (18)$$

$$\text{donde } c(\alpha) = \sqrt{-\frac{1}{2} \ln(\alpha)}, \quad (19)$$

donde n y m son los tamaños del primer y segundo vector respectivamente.

La prueba de Kolmogorov-Smirnov de dos muestras se halló con la función *kstest2* de Matlab. La decisión de la prueba es 1 si H_0 es rechazada con un nivel de significancia del 5%, y 0 en caso contrario, es decir, que las variables comparadas sí provienen de la misma distribución continua.

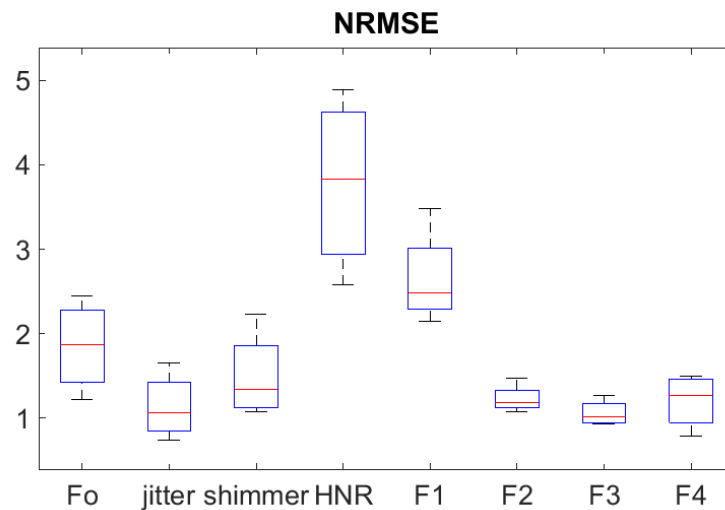
El valor de H_0 fue de 1 para todas las medias acústicas comparadas en el presente trabajo de grado, por lo tanto se descarta que provienen de la misma función de densidad de probabilidad.

⁸⁹ CERQUEIRA, D. A Test of Hypotheses for Random Graph Distributions Built from EEG Data. IEEE Transactions on Network Science and Engineering. 2017, vol. 4, nº 2, p. 77.

⁹⁰ *Ibid.*, p. 77.

Normalized root mean square error (NRMSE): También se halló el normalized root mean square error (NRMSE) entre las medidas acústicas provenientes de las frases de audio limpias y de las frases de audio con ruido de cada hablante. Luego se calculó un promedio total por medida acústica. Esta medida permitió comparar las diferentes medidas acústicas entre sí y observar cuales se ven más afectadas por el ruido de MRI y cuales no tanto.

Figura 27. Valor promedio NRMSE de cada medida acústica.



En la Figura 27 se observa que el error es mayor para el HNR y el F1 y las medidas que menos se ven afectadas son el jitter, el F2 y el F3.

CONCLUSIONES:

En el presente trabajo se evaluó la influencia de sistemas de adquisición de imágenes de resonancia magnética (MRI) sobre parámetros acústicos de la señal de voz. Dentro de las medidas acústicas se da preferencia a aquellas pertenecientes a la categoría de medidas relacionadas con la periodicidad y la envolvente espectral de la señal de voz., tales como los coeficientes cepstrales de la frecuencia de mel (MFCC), los coeficientes de codificación de predicción lineal (LPC), y otras medidas acústicas como: el jitter, shimmer, frecuencia fundamental (F0), los formantes y relación armónico a ruido (HNR). En particular, el análisis de la influencia del ruido MRI se realiza sobre la base de datos USC-TIMIT, la cual contiene videos de MRI de la sección sagital del tracto vocal en paralelo con registros de audio de habla continua, en el idioma inglés de 10 hablantes, pero solo cuenta con datos de articulografía electromagnética (EMA, *Electromagnetic Articulograph*) para cinco de esos hablantes, sin embargo en el presente proyecto de grado se trabajó con 4 de esos 5 hablantes. En general, la influencia es perceptible y los parámetros acústicos se ven afectados considerablemente. Los más afectados son el HNR y el primer formante.

Para medidas rms por encima de los 3.5 dB los cambios en la señal de voz ya serán perceptibles para algún oyente con un sistema auditivo sano⁹¹. De los resultados mostrados en la Tabla.6 se observa que para el espectro de LPC la diferencia es de 1.2 veces el valor de cambios ya perceptibles; para MFCC es en promedio 1.6 veces; y el de LTAS es de 1.5 veces el valor de cambios ya perceptibles. Por otra parte, de las comparaciones de espectros obtenidos por MFCC, se encontró que no hay una fuerte correlación entre ellos, puesto que en promedio el coeficiente de

⁹¹ A, Gray; M, John. Op. cit., p. 388.

Pearson obtenido es de 0.54; en cambio para los espectros de LPC y LTAS los coeficientes de Pearson obtenidos son en promedio de 0.75 y 0.84 respectivamente.

Por lo tanto, si se desea evitar la influencia del ruido se recomienda el uso de una representación espectral por LPC, porque además de presentar la menor distancia, cuenta con una fuerte relación entre espectro de señales limpias y con ruido, como lo comprueba su correspondiente coeficiente de Pearson.

La prueba de hipótesis F , T y de Kolmogorov-Smirnov, demostraron que las medidas acústicas (F_0 , jitter relativo, shimmer relativo, HNR y los primeros cuatros formantes), extraídas tanto de las señales de voz limpia (archivos EMA), como de las señales de voz con ruido (archivos MRI), tienen diferentes varianzas, medias y función de distribución, respectivamente. Además, con el NRMSE se identificó que las medidas acústicas que más se ven influenciadas por el ruido de MRI son el primer formante y la HNR. En cambio los que menos se ven influenciados son el jitter relativo, y el tercer formante.

Respecto a la calidad de la voz adquirida con los sistemas de MRI, los resultados muestran que la la señal de voz (SIG) está entre muy distorsionada y bastante distorsionada. El ruido debido a los dispositivos presentes en la sala de escaneo de MRI (BAK), presente aún después del filtrado, está entre muy molesto y algo molesto; y la calidad en general (OVL) es entre mala y mediocre. Asimismo, la medida PESQ de los 4 hablantes no alcanza siquiera a ser de 2, es decir, la percepción de la calidad de las señales de voz con ruido de MRI de la base de datos USCM-TIMIT no es medianamente aceptable. Además, el valor promedio de LLR para los 4 hablantes es cercano a 1, es decir, la mitad del máximo valor de distorsión, lo cual indica que el filtro de reducción de ruido no elimina satisfactoriamente ruidos parásitos presentes en los segmentos de voz de las señales de audio.

De lo anterior se puede decir que el desarrollo de métodos que permitan reducir el ruido de MRI con niveles de aceptación perceptual satisfactorio es aun campo que requiere mayor investigación.

Finalmente, para trabajos futuros se podrían desarrollar técnicas que permitan eliminar el efecto de reverberación o analizar la influencia de la posición del paciente dentro del escáner de MRI, sobre los parámetros descriptores de la voz, puesto que estos influyen sobre el rendimiento de las técnicas de cancelación de ruido desarrolladas.

BIBLIOGRAFÍA:

A,Gray; M,John.Distance Measures for Speech Processing. IEEE. 1976,vol. 4, nº 5, pp. 380-390.

ALÍAS, F; SOCORÓ, J y SEVILLANO, X. A review of physical and perceptual feature extraction techniques for speech, music and enviromental sounds. Applied Sciences, 2016, pp. 350-353.

BAO,H; PANAHI, I y BRIGGS, R. Blind dereverberation for fMRI noise based on SIMO linear prediction method. IEEE. 2007, pp. 2831-2834.

BOERSMA, P. Accurate short term analysis of the fundamental frequency and the harmonics to noise ratio of a sampled sound. IFA Proceedings, 1993.

BOERSMA, P. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound.Proceedings of the Institute of Phonetic Sciences. 1993, nº 17, pp. 97-110.

BRESCH,E ,et al. Synchronized and noise-robust audio recordings during realtime magnetic resonance imaging scans. J Acoustic Soc Am. 2006, vol. 120, no. 4, pp. 1791-1794.

CERQUEIRA, D. A Test of Hypotheses for Random Graph Distributions Built from EEG Data. IEEE Transactions on Network Science and Engineering. 2017,vol. 4, nº 2, pp. 75-82.

CHATAN, V; CHANDAN, R e ISSA, P. Improving Quality and Intelligibility of Speech using Single Microphone for the Broadband fMRI Noise at Low SNR. IEEE. 2016, pp. 3674-3678.

E. K. Introductory Functional Analysis with Applications. John Wiley & Sons, 1989.

ENGWALL, O. A Revisit to the Application of MRI to the Analysis of Speech Production-Testing our Assumptions. Sydney, 2003.

FARRÚS, M; HERNANDEZ, J y EJARQUE, P. Jitter and shimmer measurements for speaker recognition. In proceedings of 8th annual conference of the international speech communication association (InterSpeech). 2007, pp. 778-781.

HAY MAR, S, et al. Comparison of Three Auditory Frequency Scales in Feature Extraction on Myanmar Digits Recognition. 2018 10th International Conference on Information Technology and Electrical Engineering(ICITEE), Kuta, 2018, pp. 150-156.

HOITING, G. Measuring MRI noise. Groninga, 2005.

HU, Y; LOIZOU, P. Evaluation of Objective Quality Measures. IEEE Transactions on audio, speech, and language processing. Enero, 2008, vol. 16, nº 1, pp. 229-238.

HUANG, X; ACERO, A y HON, H. Spoken Language Processing: A guide to theory, algorithm, and system development, Prentice Hall, pp. 1-931.

I.-T. R. p.835. Subjetive Test Methodology for Evaluating Speech Communication Systems that Include Noise Suppression Algorithm. 2003.

ITU_T Recommendation, P.862, 2001.Perceptual Evaluation of Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs, 2001.

KHAN, I. MRI Scanner`s Vibration Isolation: Experimental Measurements,Analysis Techniques and Analytical Models. 31 January 2017,pp. 350-353.

KOLOMOGOROV, A; FOMIN,S. Introductory Real Analysis. Dover Publications, 1970.

LOIZOU, P. Speech Enhancement: theory and Practice, Boca Raton, 2007.

MACKIEWICH, B. A Thesis Subnitted in Partial Fulfillment of the Requirements for the Degree of Master of Science,» de Intracranial Boundary Detection and Radio Frecuency Correction in Magnetic Resonance Images, Simon Fraser University, 1995.

MECHEFSKE, C, et al. Acoustic noise reduction in a 4 T MRI scanner. MAGMA. 2002, vol. 13, pp. 172-176.

MONTAZERI, V; PATHAK, N y PANAHI, I.Two-channel multi-stage speech enhancement for noisy fMRI enviroment. IEEE.2013, vol. 36, nº 2, pp. 60-70.

NARAYANAN, S, et al. An approach to real-time magnetic resonance imaging for speech production. Acoustical Society of America. 2004, pp. 1771-1776.

NARAYANAN,S ,et al. USC-TIMIT: A database of multimodal speech production data. Acoustical Society of America, 2014,pp. 10-22.

NESSAIVER, M,et al. Recording High Quality Speech During Tagged Cine-MRI Studies Using a Fiber Optic Microphone. Journal of Magnetic Resonance Imaging. 2006, pp. 92-97.

PATINO SAUCEDO, A; SEPULVEDA SEPULVEDA, A y GÓMEZ CAJAS, D. Phoneme recognition system using articulatory-type information.TECCIENCIA. 2015, vol. 10, pp. 11-14.

PRESS, W, et al. Numerical Recipes in C: The art of scientific computing, Cambridge University Press, 1992, pp. 1-994.

PRIBIL, J; PRIBILOVA, A y FROLLO,I. Vibration and Noise in Magnetic Resonance Imaging of the Vocal Tract: Differences between Whole-Body and Open-Air Devices. Sensors.5 Abril 2018, pp. 350-353

PRIBIL,J; PRIBILOVA,A y FROLLO, I. Automatic GMM-Based evaluation of noise suppression in the speech signal recorded during phonation in the open-air MRI. De 2017 40th International Conference on Telecommunications and Signal Processing (TSP).Barcelona, 2017, vol. 17, nº 2, pp. 257-263.

RAMANARAYANAN, V, et al. Analysis of Speech Production Real-Time MRI. Computer Speech & Language. 2018, pp. 220-230.

RASHID, N; KIBRIA, F. MRI SCAN. Weebly. [En línea]. Disponible en: <https://snc2dmri.weebly.com/components--functions.html>.

RONDINONI, C. Effect of scanner acoustic background noise on strict resting-state fMRI. Brazilian Journal of Medical and Biological Research. 2013, pp. 359-367.

SCHWARZ, D; RODET, X. Spectral Envelope Estimation and Representation for Sound Analysis Synthesis. IRCAM, 1998.

SCOTT, A. Speech MRI: Morphology and function. Physica Medica. 2014, vol. 30, pp. 604-618.

SEPULVEDA, Alexander. Estimation of Articulatory Parameters from the Acoustic Speech Signal. Manizales, 2012, pp. 11-20.

SEPULVEDA, A; CASAS, D y CASTELLANOS, G. Importancia de las frecuencias de resonancia del tracto vocal en la estimación de posiciones articulatorias. Revista Ingeniería Biomédica. 2012, vol. 6, nº 11, pp. 53-60.

SEPULVEDA, Alexander; CASTELLANOS, G. Estimación de la frecuencia fundamental de señales de voz usando transformada Wavelet. Scientia et Technica. 2004, nº 24, pp. 7-11.

SHARMA, H. MRI physics-basic principles. Acta Neuropsychiatrica. Agosto 2009, pp. 200-201.

SPIEGEL, M. Theory and problems of Statistics in SI units. Schaum's Outline Series, 1972.

STEGGER, I , et al. Cardiac PET/MRI. Curr Cardiovasc Imaging Rep. Apr 2013, vol. 6, no. 2, pp. 169-178.

STONE, M. Comparison of speech production in upright and supine position. Acoustical Society of America. 2007, pp. 532-541.

ANEXOS

ANEXOS A.

Frases de la base de datos USC-TIMIT

Tabla 7. Las primeras 100 frases que conforman la base de datos USC-TIMIT

This was easy for us.	Don't do Charlie's dirty dishes.	Any contributions will be greatly appreciated.	The misprint provoked an immediate disclaimer.
Jane may earn more money by working hard.	Help celebrate your brother's success.	Continental drift is a geological theory.	A large household needs lots of appliances.
She is thinner than I am.	Only the most accomplished artists obtain popularity.	Regular attendance is seldom required.	Cut a small corner off each edge.
Bright sunshine shimmers on the ocean.	Critical equipment needs proper maintenance.	Challenge each general's intelligence.	Iguanas and alligators are tropical reptiles.
Nothing is as offensive as innocence.	Young people participate in athletic activities.	We got drenched from the uninterrupted rain.	Masquerade parties tax one's imagination.
Why yell or worry over silly items.	Barb's gold bracelet was a graduation present.	Last year's gas shortage caused steep price increases.	Penguins live near the icy antarctic.
Where were you while we were away?	Stimulating discussions keep students' attention.	Upgrade your status to reflect your wealth.	Guess the question from the answer.
Are your grades higher or lower than Nancy's.	Etiquette mandates compliance with existing regulations.	Eat your raisins outdoors on the porch steps.	Medieval society was based on hierarchies.
He will allow a rare lie.	Biblical scholars argue history.	Porcupines resemble sea urchins.	Project development was proceeding too slowly.
Will robin wear a yellow lily.	Elderly people are often excluded.	Spring street is straight ahead.	Kindergarten children decorate their classrooms for all holidays.
Swing your arm as high as you can.	Basketball can be an entertaining sport.	Cliff's display was misplaced on the screen.	Special task forces rescue hostages from kidnappers.
Before Thursday's exam review every formula.	Addition and subtraction are learned skills.	An official deadline cannot be postponed.	Call an ambulance for medical assistance.

The museum hires musicians every evening.	Grandmother outgrew her upbringing in petticoats.	Fill that canteen with fresh spring water.	He stole a dime from a beggar.
A roll of wire lay near the wall.	At twilight on the twelfth day we'll have chablis.	Gently place Jim's foam sculpture in the box.	You must explicitly delete files.
Carl lives in a lively home.	Catastrophic economic cutbacks neglect the poor.	Bagpipes and bongos are musical instruments.	A huge tapestry hung in her hallway.
Alimony harms a divorced man's wealth.	Ambidextrous pickpockets accomplish more.	Doctors prescribe drugs too freely.	Birthday parties have cupcakes and ice cream.
She wore warm fleecy woolen overalls.	Even a simple vocabulary contains symbols.	Will you please describe the idiotic predicament.	His scalp was blistered from today's hot sun.
Alfalfa is healthy for you.	The eastern coast is a place for pure pleasure and excitement.	It's impossible to deal with bureaucracy.	She slipped and sprained her ankle on the steep slope.
When all else fails use force.	The lack of heat compounded the tenant's grievances.	Good service should be rewarded by big tips.	The best way to learn is to solve extra problems.
Although always alone we survive.	Academic aptitude guarantees your diploma.	My instructions desperately need updating.	His sudden departure shocked the cast.
Only lawyers love millionaires.	The prowler wore a ski mask for disguise.	Cooperation along with understanding alleviate dispute.	Tugboats are capable of hauling huge loads.
Did dad do academic bidding?	We experience distress and frustration obtaining our degrees.	Cement is measured in cubic yards.	A muscular abdomen is good for your back.
Help greg to pick a peck of potatoes.	The singer's finger had a splinter.	Primitive tribes have an upbeat attitude.	The cartoon features a muskrat and a tadpole.
A good attitude is unbeatable.	The legislature met to judge the state of public education.	Flying standby can be practical if you want to save money.	The emblem depicts the acropolis all aglow.
Coconut cream pie makes a nice dessert.	Chocolate and roses never fail as a romantic gift.	'The thinker' is a famous sculpture.	Clasp the screw in your left hand.