

# Simulador de Deficiencias Auditivas implementado en MATLAB

A. Uriz<sup>1</sup>, P. Agüero<sup>2</sup>, F. Denk<sup>2</sup>, J. Castiñeira<sup>1</sup>, J.C. Tulli<sup>2</sup>, J.M. Garin<sup>2</sup> y G. Aranda<sup>2</sup>

<sup>1</sup>CONICET, UNMdP, Laboratorio de Comunicaciones, Mar del Plata Argentina, J. B. Justo 4302, +5402234816600.  
ajuriz@fi.mdp.edu.ar, casti@fi.mdp.edu.ar

<sup>2</sup>UNMdP, Laboratorio de Comunicaciones, Mar del Plata, Argentina, J. B. Justo 4302, +5402234816600.  
pdaguero@fi.mdp.edu.ar, franciscodenk@fi.mdp.edu.ar, jctulli@fi.mdp.edu.ar, jmgarin@fi.mdp.edu.ar, garanda@fi.mdp.edu.ar

**Resumen.** El estándar para compresión de audio en alta calidad MPEG, es una norma desarrollada con el fin de comprimir señales sin perder información relevante. En particular, el estándar MPEG-1 Audio Layer III se basa en el modelo Psicoacústico I. A lo largo del presente trabajo se desarrolla un sistema que explota las características de dicho modelo para simular distintos tipos de discapacidades auditivas partiendo de las curvas que representan el umbral de audición del oído en función de la frecuencia.

**Palabras clave:** Discapacidad auditiva, audiometría, Procesamiento Digital de Señales, Compresión de Voz, Codificación de Voz.

## 1. Introducción.

En nuestra Sociedad existe un gran número de personas con algún grado de discapacidad auditiva (en Argentina, según el Censo del 2001, alrededor de 300.000). Es por ello que se vuelve necesario producir sistemas de asistencia a las mismas [1]. Uno de los problemas que se encuentran los desarrolladores de estos sistemas es que para el proceso de experimentación es necesario recurrir a evaluadores humanos. Esto trae aparejado dos problemas: uno es que se necesita disponer de una cierta cantidad de evaluadores voluntarios dispuestos a colaborar en tiempo y forma, pero hay ocasiones en que es difícil encontrar personas con una determinada discapacidad auditiva. En este contexto se desarrolla el presente trabajo. De este trabajo se desprenden también otras utilidades, ya que con una herramienta de este tipo se puede conocer *cómo percibe el mundo una persona con una discapacidad auditiva*. De este modo se pueden realizar consideraciones especiales a la hora de diseñar determinadas señales de alarma sonoras. Es decir, esta herramienta más allá de ser un soporte para desarrollos tecnológicos experimentales, es una herramienta necesaria para realizar diseños que contemplen a toda la Sociedad por igual.

El trabajo se basa en una implementación del modelo MPEG-1 Audio Layer 3 [2,3] el cual es un estándar de compresión de audio que se basa en el modelo auditivo humano [4] y funciona descartando la información que el mismo no logra percibir. De esta forma, si se modifica el modelo auditivo de modo tal que represente al sistema de audición de una persona con una determinada discapacidad auditiva, se pueden obtener audios afectados por una dolencia en particular.

La estructura del trabajo es la siguiente: a lo largo de esta Sección se desarrollan los fundamentos teóricos del mismo, presentando las partes relevantes del modelo utilizado, pero también las deficiencias auditivas que serán objeto de estudio a lo largo del trabajo. La Sección 2 presenta el modelo implementado junto con las consideraciones realizadas a la hora de implementarlo. La Sección 3 describe los experimentos subjetivos utilizados para validar los resultados, mientras que en la Sección 4 se presentan las conclusiones del trabajo y los lineamientos a seguir en el futuro.

### 1.1. Modelo Auditorio Humano

Como se mencionó en la Introducción, el modelo propuesto trabaja en base al estándar de compresión MPEG-1 Audio Layer 3, el cual es utilizado por el algoritmo MP3. Este estándar de compresión se basa en el modelo psicoacústico I [2,3]. Un diagrama en bloques del sistema de codificación se aprecia en la Figura 1.

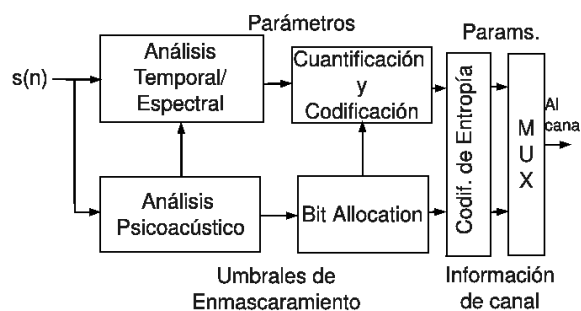


Figura 1. Modelo de un sistema genérico de codificación perceptual.

Lo relevante en este modelo es que la codificación se realiza descartando la información acústica que no es percibida por el oído humano. En base a ello, logra niveles de compresión muy elevados. Las partes más importantes del mismo se describen en las siguientes secciones.

### 1.1.1. Umbral Absoluto de Audición (ATH)

La función del Umbral Absoluto de Audición (**ATH**) representa la curva de la intensidad de presión de sonido requerida (SPL) en función de la frecuencia en un ambiente sin ruido. Es decir, es la mínima cantidad de potencia que requiere una componente de una determinada frecuencia para ser percibida por el oído de una persona joven sin ninguna discapacidad auditiva. La curva del ATH se puede apreciar en la Figura 2.

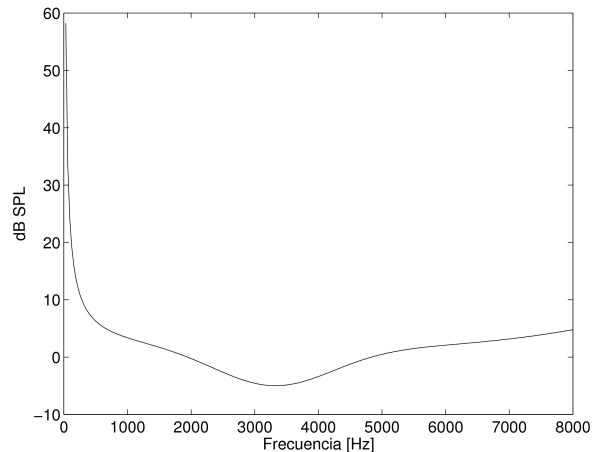


Figura 2. Umbral Absoluto de Audición para una persona joven

Cabe destacar que esta curva puede ser adaptada para el caso en que una persona sufra de una deficiencia auditiva y se verá más en detalle en la Subsección 1.2.

### 1.1.2. Bandas Críticas y Enmascaramiento

Dado que la información espectral de la señal de voz no se distribuye de forma homogénea en el dominio de la frecuencia, el mismo se divide en bandas con el fin de, trabajar cada rango de frecuencias con la resolución que corresponda, y también así reducir discontinuidades que pudieran surgir por el solapamiento entre bandas adyacentes.

El enmascaramiento es un fenómeno que ocurre cuando una componente de una determinada potencia impide que otra componente sea percibida. Existen dos tipos de enmascaramiento: el enmascaramiento simultáneo ó espectral y el enmascaramiento no simultáneo ó temporal. El primero de los dos se caracteriza por ocurrir cuando una componente (ya sea tonal o no tonal) enmascara a otra componente de menor potencia que se encuentra dentro de la misma banda de frecuencias. Este enmascaramiento se debe a la falta de selectividad de los filtros que modelan el oído humano. Además, debe tenerse en cuenta que este es un fenómeno que no es espectralmente asimétrico. En la Figura 3 se aprecian las características más importantes del mismo.

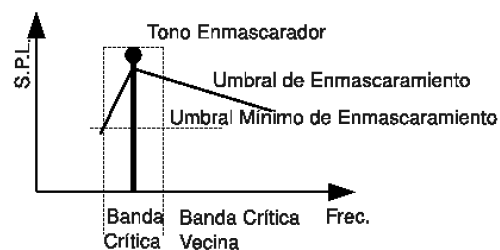


Figura 3. Parámetros que intervienen en el enmascaramiento espectral.

El segundo tipo de enmascaramiento, el temporal, ocurre en el dominio del tiempo y se debe a que el oído necesita un tiempo de recuperación desde la desaparición de una componente hasta que pueda percibir otra de menor potencia. De no cumplirse el mismo, se dice que la segunda componente es enmascarada por la primera. Este tiempo es de al menos 5ms.

## 1.2. Deficiencias auditivas a considerar en este trabajo

A lo largo de este trabajo se pretende simular una serie de discapacidades auditivas, es por ello que primero se procede a presentar cada una de ellas con un enfoque conceptual.

### 1.2.1. Oído sin discapacidades

Con el propósito de fijar una referencia para los experimentos, se implementa el modelo del oído de una persona joven, sin ninguna discapacidad. Este sistema permite a los evaluadores apreciar el efecto de procesamiento sólo del algoritmo de codificación. Esto se debe a que esta variación del sistema no afecta el sonido con ningún grado de deficiencia auditiva. La curva correspondiente a su Umbral Absoluto de Audición (ATH) se aprecia en la Figura 1.

### 1.2.2. Oído afectado por una sordera severa

Este tipo de discapacidad [6, 7, 8] se caracteriza por el hecho que la persona que la padece deja de percibir las frecuencias superiores. Generalmente es una pérdida gradual que empieza por las más altas frecuencias y se va propagando hacia las más bajas. Es un fenómeno asociado con el envejecimiento de oído, el cual se modela sumando un rechazo de al menos 50dB en las frecuencias superiores a 400Hz a la curva ATH, esto provoca que los sonidos que tengan un nivel por debajo de dicha curva no sean percibidos por el oyente. La Figura 4 muestra un modelo de esta dolencia, donde se simula una discapacidad con una frecuencia de corte de 400Hz.

### 1.2.3. Oído afectado por una discapacidad auditiva sensorineural

Para este tipo de discapacidad [9], la curva ATH presenta la propiedad de que el umbral de percepción se incrementa en función de la frecuencia. Por este motivo la persona aquejada por esta dolencia experimenta una degradación progresiva en la percepción de frecuencias superiores. La curva del ATH para este tipo de dolencia se representa utilizando línea rayada en la Figura 5.

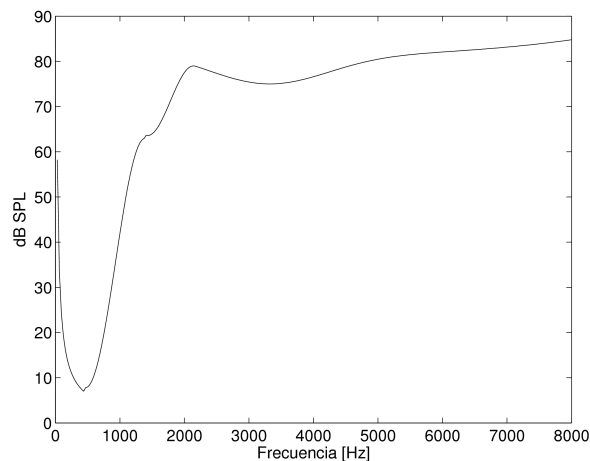


Figura 4. Curva SPL en función de la frecuencia para una discapacidad auditiva severa.

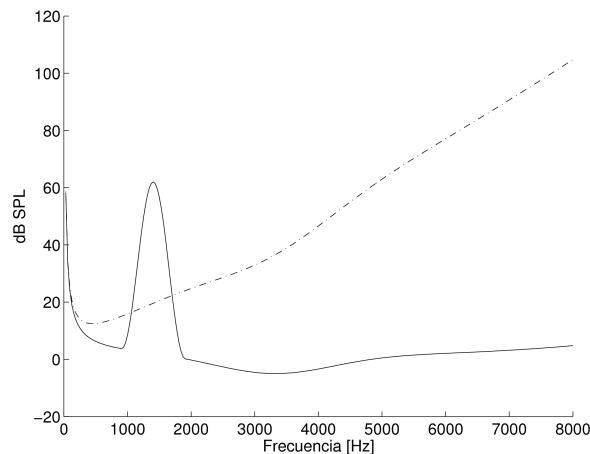


Figura 5. Función SPL en función de la frecuencia para una discapacidad sensorineural y una de origen traumático.

### 1.2.4. Oído afectado por una discapacidad de origen traumático

Esta dolencia [7, 8, 10] se origina al exponer a una persona a un sonido constante de gran intensidad, por ejemplo la maquinaria de un ambiente industrial. Esto se traduce en una curva ATH con un rechazo abrupto a una banda de frecuencias específica (por ejemplo un rechazo de 50 dB entre  $f=1000\text{Hz}$  y  $f=3000\text{Hz}$ ). En la Figura 5, utilizando trazo continuo se representa la curva del ATH correspondiente a este tipo de dolencia.

## 2. Metodología

### 2.1. Sistema Propuesto

A lo largo de esta Sección se presenta el modelo desarrollado con el fin de implementar las discapacidades auditivas presentadas en la Sección 1.2. Con este propósito se genera un sistema que se divide en dos etapas. La primera toma la señal de audio, la procesa con el modelo perceptual presentado en la Sección 1.1, y como resultado se obtiene la información que debe ser utilizada para simular la discapacidad, descartando la información que no sea percibida por el modelo en particular bajo estudio. Cabe destacar que si a este modelo se le modifica la curva del umbral absoluto de audición (ATH) de modo que simule el de un oído con alguna de las discapacidades presentadas en la Sección 1.2, a la salida de este bloque se obtiene como resultado la información que es descartada y cual debe ser tomada en cuenta a la hora de resintetizar el audio, lo que depende exclusivamente de la dolencia bajo estudio.

La segunda etapa del sistema, se encarga de tomar la señal de audio, descomponerla utilizando segmentos de 32ms de longitud (512 muestras, con una frecuencia de muestreo de 16KHz), aplicarles la Transformada Rápida de Fourier (FFT, por sus siglas en inglés). En esta instancia, teniendo en cuenta la información proveniente del modelo auditorio, se realiza la selección de la información espectral que es percibida por el oído de la persona, en el caso particular bajo estudio. De este modo, toda la información que el modelo auditorio considera que no es percibida es eliminada.

Luego de este proceso, se debe realizar un ajuste entre los segmentos con el fin de reducir las discontinuidades en las fronteras de los segmentos, las cuales se originan principalmente por tres motivos: el **leakage**[13,15], el **efecto de estacas**[13,15] y el efecto debido a los cambios de amplitud de las componentes entre segmentos consecutivos. El primero de los efectos, se debe a que al realizar el segmentado se asume una periodicidad en la señal, lo cual no es válido. En segundo lugar, el efecto de estacas se debe a que el segmento de señal procesado no abarca un número entero de períodos de la misma, cabe destacar que como se dijo previamente, la señal de entrada no es periódica, pero dentro de cada segmento de análisis, puede asumirse la periodicidad de la misma. Si se considera que la señal de voz tiene componentes de infinitas frecuencias, se hace evidente la existencia de este efecto, el cual se traduce en la separación de cada componente en otras de frecuencias adyacentes a la obtenida. Esto se aprecia en la Figura 6, donde  $f_{IN}$  representada en línea gruesa a rayas, es la componente de entrada, la cual no coincide con ninguna de las componentes discretas. Por este motivo se generan componentes de amplitud decreciente en las divisiones  $(1, \dots, k-2, k-1, k, k+1, k+3, \dots, N)$  las cuales se representan utilizando una línea fina de trazo continuo. El problema que aparece al procesar este segmento utilizando el modelo psicoacústico I es que debido al enmascaramiento espectral, se eliminan todas las componentes, a excepción de la componente de  $k$ -ésima, la más próxima al tono afectado. En la Figura 6, esta componente se representa en trazo grueso continuo.

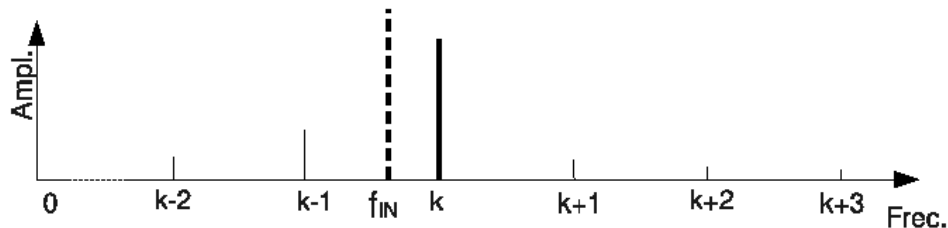


Figura 6. Error cometido por el efecto de estacas.

El problema es que en este caso el enmascaramiento elimina información requerida para la resíntesis del audio, es por ello que con el fin de reducir el error debido a este efecto, una vez procesado el segmento, se regeneran las dos componentes más próximas al tono afectado por el efecto de estacas, lo cual se aprecia en la Figura 7. Esto genera un aumento notable en la calidad del audio sintetizado.

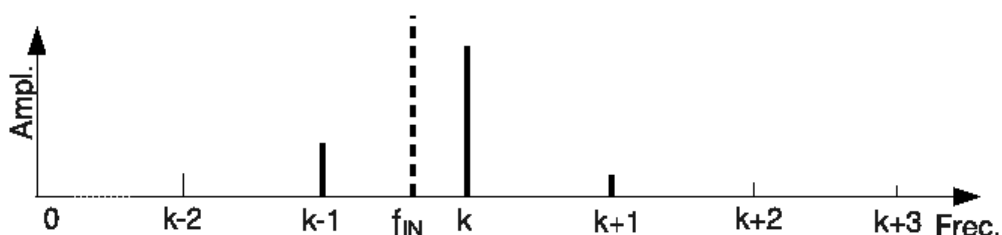


Figura 7. Corrección Implementada.

Cabe destacar que esta corrección sólo contempla el efecto de estacas. Dado que en las condiciones de trabajo propuestas, es posible asumir la periodicidad de las componentes de mayor energía de la señal, para las cuales el error por leakage es mucho menor que el debido al efecto de estacas. Esto se debe fundamentalmente a que el ancho del segmento  $N$  utilizado (512 muestras) si bien es lo suficientemente pequeño como para asegurar la periodicidad de las componentes, brinda una resolución espectral de 31.25Hz, con una frecuencia de muestreo de 16KHz.

Otro efecto que debe considerarse y que impacta en el análisis es que si bien se asume estacionariedad en el análisis de señales de voz, esto no es así en la realidad. De hecho, al hacer un análisis entre segmentos consecutivos se aprecia que la amplitud de las componentes de cada uno de ellos varía notablemente. Este efecto produce discontinuidades en la energía del audio sintetizado, las cuales impactan en su calidad. Por ello, se redujo la separación entre segmentos, aumentando el solapamiento entre los mismos con el fin de asegurar que no existan grandes diferencias entre las amplitudes de las componentes de dos segmentos consecutivos. Dado que este proceso aumenta notablemente el costo computacional del sistema, se optó por utilizar un solapamiento que si bien incrementa la calidad del audio, no afecta significativamente el tiempo de procesamiento de los audios.

Por último, el espectro resultante de las modificaciones previamente desarrolladas es convertido al dominio del tiempo mediante una Transformada Rápida de Fourier Inversa (IFFT) los cuales son sintetizados mediante el método de la suma solapada en el dominio del tiempo (*Time Domain Overlap and Add* ó **TD-OLA**)[15]. Un diagrama en bloques del sistema completo se aprecia en la Figura 7, donde el bloque llamado “**sistema articulatorio**” representa las consideraciones realizadas para reducir los errores por efecto de estacas y por las variaciones de la señal.

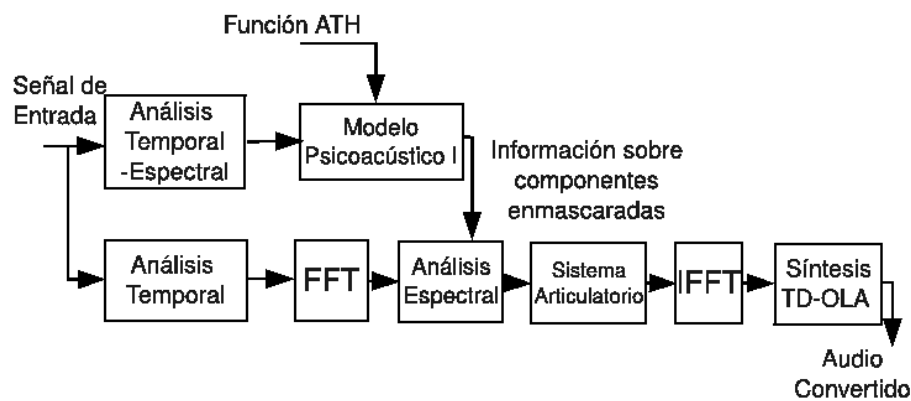


Figura 8. Diagrama de bloques del sistema propuesto.

### 3. Resultados y Discusión

Con el fin de analizar el funcionamiento del sistema se realizaron experimentos con una base de datos de 35 oraciones en idioma español peninsular, pronunciadas por dos locutores masculinos y dos locutores femeninos. La duración promedio de las oraciones fue de 4 segundos.

Se realizaron dos experimentos subjetivos, ambos utilizando el sistema propuesto en la Sección anterior, el cual corrige los errores producidos por el efecto de estacas y los del modelado fonético. 15 voluntarios evaluaron una serie de audios tanto en **calidad** como en **inteligibilidad**, calificando a cada una con puntajes de 1 a 5, según el siguiente criterio: 1- (Calidad/Inteligibilidad) pésima a 5- (Calidad/Inteligibilidad) excelente. Con el fin de analizar el funcionamiento del sistema se tomaron como referencia audios sin ningún tipo de procesamiento. El experimento está compuesto por un conjunto de audios afectados por las siguientes discapacidades simuladas:

- **Audios sin ningún tipo de dolencia.** Para la simulación es utilizada una curva ATH como la que se muestra en la Figura 2.
- **Audios afectados por una discapacidad auditiva severa.** Se utiliza la curva ATH presentada en la Figura 4.
- **Audios afectados por una discapacidad sensorineural.** Se utiliza la curva ATH graficada en trazo rayado en la Figura 5.
- **Audios afectados por una dolencia de tipo traumático.** Se utiliza la curva ATH graficada en trazo continuo en la Figura 5.

#### 3.1. Calidad

El primer experimento se realiza para medir la calidad de los audios procesados. Este experimento resulta fundamental ya que permite validar el experimento de inteligibilidad, ya que si se obtienen resultados elevados ( en particular el caso del oído simulado sin discapacidad) significa que el algoritmo de procesamiento no afecta la señal y por lo tanto el único efecto que sufren los audios es la discapacidad simulada. Los resultados se presentan en la Tabla 1. En la misma se puede apreciar que la calidad de todos los audios es relativamente alta y del mismo orden para ambos géneros de locutores. Por este motivo, es posible realizar el siguiente experimento referente a la inteligibilidad.

	Resultados Loc. Masculinos	Resultados Loc. Femeninos	Resultados Conjunto
<b>Discapacidad Severa</b>	3.00	3.20	3.10
<b>Discapacidad Sensorineural</b>	3.20	3.90	3.55
<b>Discapacidad Traumática</b>	3.00	2.70	2.85
<b>Sin Discapacidad</b>	3.80	3.70	3.75
<b>Referencia</b>	5.00	4.80	4.90

Tabla 1. Resultados del experimento que evalúa la calidad de los audios procesados.

### 3.2. Inteligibilidad

Este experimento analiza los problemas en la inteligibilidad de cada uno de los audios para cada una de las condiciones de discapacidad simulada.

	Resultados Loc. Masculinos	Resultados Loc. Femeninos	Resultados Conjunto
<b>Discapacidad Severa</b>	1.50	1.10	1.30
<b>Discapacidad Sensorineural</b>	3.00	2.60	2.80
<b>Discapacidad Traumática</b>	4.00	3.50	3.75
<b>Sin Discapacidad</b>	5.00	5.00	5.00
<b>Referencia</b>	5.00	5.00	5.00

Tabla 2. Resultados del experimento que evalúa la inteligibilidad de los audios procesados.

Los resultados son presentados en la Tabla 2, donde se puede apreciar que los audios que obtienen los resultados mas elevados, son los que no fueron sometidos a ningún tipo de discapacidad simulada, tal como era de esperarse. En segundo lugar, aparecen los audios afectados por una discapacidad de origen traumática. Esto es coherente, ya que este tipo de discapacidad afecta a algunas componentes, permaneciendo las restantes sin modificaciones. Para el caso de una dolencia del tipo sensorineural. El caso es distinto, ya que dado que la curva ATH crece linealmente con la frecuencia, el rechazo a las componentes de frecuencias superiores aumenta de la misma forma, por este motivo los puntajes obtenidos por los audios correspondientes son bajos. En último lugar, aparecen los audios procesados simulando una sordera grave, la cual elimina toda la información de las componentes por encima de los 350Hz, se puede apreciar que el resultado del experimento en este caso es muy bajo.

Un aspecto que se debe considerar es que para todos los audios pronunciados por locutores masculinos el puntaje del test de inteligibilidad es superior al de los audios pronunciados por mujeres. Esto se fundamenta en el hecho de que los hombres tienen sus formantes en frecuencias mas bajas que las mujeres, y por ello la información que se afecta por la sordera es menor que en el caso de un audio pronunciado por una mujer.

## 4. Conclusiones

En este trabajo se presenta un sistema que permite simular diversos tipos de discapacidades auditivas. La gran ventaja de este sistema es que la discapacidad puede ser modelada fácilmente con sólo modificar la curva que representa el Umbral Absoluto de Audición en función de la frecuencia, por una que represente dicha función. Pero correspondiente a la patología que se desee modelar.

Por otro lado, se han utilizado técnicas de procesamiento digital de señales con el fin de optimizar el rendimiento del sistema en calidad, lo cual permitió reducir notablemente la cantidad de discontinuidades, aumentando por consiguiente la calidad de los audios sintetizados.

Los trabajos futuros apuntarán a continuar mejorando la calidad de los audios sintetizados, con el fin de poder utilizar esta herramienta para el desarrollo de dispositivos de asistencia auditiva, de modo tal de reducir lo mas posible la necesidad de evaluadores.

## Referencias

- [1] A. Uriz, P. Agüero, J.C. Tulli, E. Gonzalez, F. Denk. Desarrollo de un Sistema de Compresión de Voz portátil para Pacientes Discapacitados. Proceedings of SABI 2009, 2009.
- [2] ISO/IEC JTC1/SC29/WG111 MPEG IS13813(1994). Generic Coding of Moving Pictures and Associated Audio. (1994).
- [3] T. Painter, A. Spanias. Perceptual Coding of Digital Audio. Proceedings of the IEEE. Vol.88 no.4, pp. 451-513, 2000.
- [4] M. Slaney. Auditory toolbox. Technical Report 010, 1998.
- [5] H. Fletcher. Auditory patterns Rev. Mod. Phys. pp. 47-65, 1940.
- [6] E. Vilchur. Signal processing to improve speech intelligibility in perceptive deafness. Journal of the Acoustical Society of America. Vol.53, no.6, pp. 1646-1657, 1973.
- [7] T. Baer, B.C. Moore, K. Kluk. Effects of low pass filtering on the intelligibility of speech in noise for people with and without dead regions at high frequencies. Journal of the Acoustical Society of America. Vol.112,no.3, pp. 1133-1144, 2002.
- [8] O. Timms. Speech Processing strategies based on the sinusoidal speech model for the profoundly Hearing Impaired. Ph.D. Thesis., 2003.
- [9] D.W. Kim, Y.C. Park, W.K. Kim, S.J. Park, W. Doh, S.W. Shin, D.H. Youn. Simulation of hearing impairment with sensorineural hearing loss. Proceedings of the 19th International Conference IEEE/EMBS, pp. 1986-1989, 1997.
- [10] J. Calupper, H. Fastl. Simulation of hearing impairment based on the Fourier Time Transform. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 857-860, 2000.
- [11] D. Pan. A Tutorial on MPEG/Audio Compression. IEEE Multimedia. pp. 60-74, 1995.
- [12] R. Arelhi, D.R.Campbell. A MATLAB Simulink Implementation of Speech Masking Based on the MPEG Psychoacoustic Model I. Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis. pp. 543-548, 2003.
- [13] J.G. Proakis, D.G. Manolakis. Digital Signal Processing: Principles, algorithms and applications. 1996.
- [14] A.V. Oppenheim, R.W. Schafer, J.R. Buck. Discrete-time signal processing. 1999.
- [15] X. Huang, A. Acero, H.W. Hon. Spoken Language Processing. A Guide of Theory, Algorithm, and System Development. 2001.
- [16] G. Fant. Acoustic Theory of Speech Production: The Hague. NL, Mouton, 1970.
- [18] E. Zwicker, G. Flottorp, S.S. Stevens. Critical Band Width in Loudness Simmation. Journal of the Acoustical Society of America. Vol. 29,no.3, pp. 548-557, 1957.