

## Simulador de Deficiencias Auditivas

A. Uriz \* P. Agüero \*\* F. Denk \*\* J. Tulli \*\* E. González \*\* J. Castiñeira \*

\* CONICET, Laboratorio de Comunicaciones, Facultad de Ingeniería,  
Universidad Nacional de Mar del Plata, Juan B. Justo 4302, 7600 Mar del  
Plata, Argentina (e-mail: [ajuriz@conicet.gov.ar](mailto:ajuriz@conicet.gov.ar), [casti@fi.mdp.edu.ar](mailto:casti@fi.mdp.edu.ar))

\*\* Laboratorio de Comunicaciones, Facultad de Ingeniería, Universidad  
Nacional de Mar del Plata, Juan B. Justo 4302, 7600 Mar del Plata,  
Argentina (e-mail: ([@fi.mdp.edu.ar](mailto:pdaguero.franciscodenk,jctulli,elgonzal)))

**Resumen** El estándar para compresión de audio de 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 perceptual para simular distintos tipos de discapacidades auditivas partiendo de las curvas que representan el umbral de audición del oído en función de la frecuencia. Se realizan experimentos objetivos y subjetivos con el fin de analizar el funcionamiento del simulador con algunos tipos de deficiencias auditivas. Copyright © 2011 CEA

**Palabras Clave:** Deficiencias Auditivas, Procesamiento Digital de Señales, MPEG-1 Audio Layer III, Modelo Psicoacústico I, Audiometría.

### 1. INTRODUCCIÓN

En la Sociedad Argentina existen alrededor de 300.000 personas con algún grado de discapacidad auditiva. Es por ello que se vuelve necesario producir sistemas de bajo costo de asistencia a las mismas. Uno de los problemas que encuentran los desarrolladores de estos sistemas es que para el proceso de experimentación es fundamental recurrir a evaluadores humanos. Esto trae aparejado dos dificultades: una es que se necesita disponer de una cierta cantidad de evaluadores voluntarios dispuestos a colaborar en tiempo y forma, y la otra es que es difícil encontrar personas con una determinada discapacidad auditiva.

Con el fin de mejorar la calidad de vida de personas con discapacidades auditivas se han desarrollado aplicaciones de asistencia tecnológica que funcionan acondicionando la señal que ingresa al oído de modo tal que sea percibida por el usuario. Es posible encontrar diversos tipos de dispositivos de asistencia, pero los más comunes son: audífonos basados en amplificadores y los más modernos que trabajan con procesamiento digital de señales. Los primeros simplemente amplifican la señal que ingresa al oído por completo o sólo la banda de frecuencias que presenta alguna dolencia, es un ejemplo de ellos Widex (2003). Por otro lado, los que trabajan con procesamiento digital de señales (Timms, 2003), (Widex, 2004), (Uriz *et al.*, 2009), (Denk *et al.*, 2010), se desarrollaron con el surgimiento de microcontroladores más veloces y de mayor capacidad. Estos permiten manipular la señal de una manera más compleja, ya que por ejemplo permiten desplazar las componentes de la señal que se encuentran en una región del espectro donde el oído presente alguna discapacidad, hacia una región donde el mismo funcione correctamente. Para ello se utilizan procesadores digital de señales (*digital signal processor* o **DSP**), los cuales combinan una gran velocidad de cómputo con capacidad de memoria de programa y periféricos de entrada y salida. Cabe destacar que estos dispositivos no sólo están orientados

a mejorar la inteligibilidad de las señales de voz, ya que están destinados a aumentar la autonomía de las personas que padecen algún tipo de dolencia auditiva. Es por ello que deben favorecer además la inteligibilidad de otras señales que son vitales para la independencia una persona. Ejemplos de estos sonidos críticos que deben ser percibidos son: bocinas de automóviles, advertencias sonoras de pasos a nivel ferroviarios y alarmas de servicios públicos, entre otros.

Todos los dispositivos de asistencia auditiva requieren en su etapa de desarrollo de experimentos que validen o refuten el funcionamiento de los mismos. Es por ello que es necesario utilizar evaluadores que analicen el funcionamiento de dichos sistemas. Dichos evaluadores deben poseer según el tipo de dispositivo que se esté desarrollando un tipo y grado de dolencia en particular, de lo que en muchos casos es difícil o imposible disponer. Por este motivo, se plantea el desarrollo de un dispositivo que permita simular el comportamiento de un oído con una dolencia específica, lo que brinda al desarrollador de dispositivos de asistencia auditiva mayor flexibilidad en su trabajo, ya que reduce al mínimo la cantidad de evaluadores humanos requeridos para el desarrollo del equipo.

El diseño de sistemas que modelen el funcionamiento de un oído con algún tipo de dolencia ha sido realizado mediante diversos enfoques. Puede encontrarse que Simpson (1991) y Kim *et al.* (1997) modelan una deficiencia del tipo sensorineural, la cual se caracteriza por ofrecer un rechazo a las frecuencias superiores linealmente creciente con la frecuencia. Por otro lado, Vilchur (1973), Baer *et al.* (2002), Timms (2003) y Muñoz *et al.* (1999) modelan discapacidades severas mediante bancos de filtros que simulan el comportamiento del oído afectado por la deficiencia. En estos trabajos, este tipo de dolencias son simuladas mediante filtros pasabajos, que presentan un rechazo a las frecuencias superiores a un cierto rango, el cual se define en función del grado de la dolencia, pero para una sordera severa es del orden de  $f = 800\text{Hz}$ . Esta aproximación de ban-

cos de filtros también ha sido utilizada en Calupper and Fastl (2000) y Paarmann and Wynne (1994) para modelar deficiencias traumáticas que pueden ser aproximadas mediante un filtro rechaza-banda, ya que se caracterizan por ser discapacidades que ofrecen un rechazo a un determinado rango de frecuencias, impidiendo que algunas componentes sean percibidas por una persona que padece una dolencia de este tipo. Si bien estos sistemas permiten aproximar el modelado de la discapacidad mediante bancos de filtros, se aprecia que se carece de una herramienta capaz de ofrecer un modelo de diversos tipos de discapacidades específicas, que presente un modelado más fidedigno de cada dolencia y además sea flexible en cuanto a su configuración.

En este contexto se desarrolla el presente trabajo, del cual 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 específica (Maniwa and Jongman, 2008), (Kates, 2010). De este modo se pueden realizar consideraciones especiales a la hora de, por ejemplo, diseñar determinadas señales de alarma sonoras. Es decir, este sistema más allá de ser un soporte para desarrollos tecnológicos experimentales tales como los descritos en Uriz *et al.* (2009) y Denk *et al.* (2010), es una herramienta necesaria para realizar diseños para una Sociedad igualitaria.

El trabajo se basa en una implementación del modelo MPEG-I Audio Layer 3 (ISO, 1994) (Painter and Spanias, 2000) el cual es un estándar de compresión de audio que se fundamenta en el modelo auditivo humano (Fletcher, 1940) (Slaney, 1998) y funciona descartando la información que el mismo no logra percibir. Por ello, al modificar el modelo auditivo de modo tal que represente al sistema de audición de una persona con una determinada deficiencia auditiva, se pueden obtener audios afectados por una dolencia en particular. Con el fin de verificar estas hipótesis se realizan experimentos objetivos y subjetivos para comprobar el funcionamiento del sistema.

La estructura del trabajo es la siguiente: a lo largo de la Sección 2 se presenta el modelo perceptual utilizado junto con las consideraciones realizadas a la hora de implementarlo. La Sección 3 desarrolla las discapacidades que serán modeladas en este trabajo. La Sección 4 describe los modelos implementados en las simulaciones. La Sección 5 presenta las condiciones en las que se realizaron los experimentos, mientras que la Sección 6 describe los experimentos utilizados para validar los resultados del simulador de deficiencias auditivas. Por último, en la Sección 7 se presentan las conclusiones del trabajo y los lineamientos a seguir en el futuro.

## 2. MODELO AUDITIVO HUMANO

Como se mencionó en la Introducción, en el modelado de deficiencias auditivas existe una variedad de trabajos que modelan estas discapacidades mediante bancos de filtros, los cuales varían sus parámetros según la discapacidad a modelar. Si bien estos sistemas son muy utilizados, su aproximación no es muy precisa ya que no están basados en modelos psicoacústicos (ISO, 1994). Por ello, con el fin de optimizar el funcionamiento de estos sistemas este trabajo pretende desarrollar un simulador de deficiencias auditivas que pueda ser configurado a partir de datos provenientes de estudios médicos (una audiometría, por ejemplo). Con el fin de realizar una simulación que modele de una forma más adecuada el fenómeno bajo estudio el modelo propuesto trabaja en base al estándar de compresión MPEG-

I Audio Layer 3, el cual es utilizado por el algoritmo MP3. Este estándar de compresión se basa en el modelo psicoacústico I (ISO, 1994) (Painter and Spanias, 2000). Un diagrama en bloques del sistema de codificación se aprecia en la Fig.1. Como puede verse, el sistema de compresión perceptual adoptado está compuesto por seis bloques principales:

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

- Sistema de Análisis temporal/espectral.
- Análisis psicoacústico.
- Asignación de Bits.
- Cuantificación y Codificación.
- Codificación por Entropía.
- Multiplexor.

De los bloques presentados previamente se utilizan sólo los de **análisis temporal-espectral** y **análisis psicoacústico**, el primero de los dos descompone la señal  $s(n)$  que ingresa al sistema en segmentos, a los cuales se les aplica una transformación mediante la transformada rápida de Fourier (FFT), con el fin de procesar la información en el dominio de la frecuencia. El segundo bloque, el de análisis psicoacústico, permite obtener la señal tal como es percibida por el oído de una persona joven en buen estado de salud auditiva, en un ambiente sin ruido. Si se opera sobre las componentes de este bloque, se modifica la señal resultante, lo que es equivalente a obtener un cambio en el modelo perceptual humano. Por lo tanto, variando de una manera adecuada estas componentes, es posible obtener la señal resultante que percibiría una persona con algún tipo de deficiencia auditiva. Las características más relevantes de este bloque son presentadas en las siguientes subsecciones.

### 2.1 Umbral Absoluto de Audición (ATH)

La función del Umbral Absoluto de Audición (*Absolute Threshold of Hearing* o **ATH**) (Fletcher, 1940) representa la curva de la intensidad de presión de sonido requerida (*Sound Pressure Level* o **SPL**) en función de la frecuencia en un ambiente sin ruido para que un sonido sea percibido. 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 ( $S(f)$ ) es descrita por (1) y la misma se puede apreciar en la Fig.2. La unidad utilizada para representar esta magnitud es el *dB*.

$$S(f) = 3.64 \left( \frac{f}{1000} \right)^{-0.8} - 6.5 e^{-0.6 \left( \frac{f}{1000} - 3.3 \right)^2} + 10^{-3} \left( \frac{f}{1000} \right)^4 \quad (1)$$

Partiendo de datos médicos obtenidos de un estudio audiométrico es posible ajustar esta curva a la correspondiente de un oído

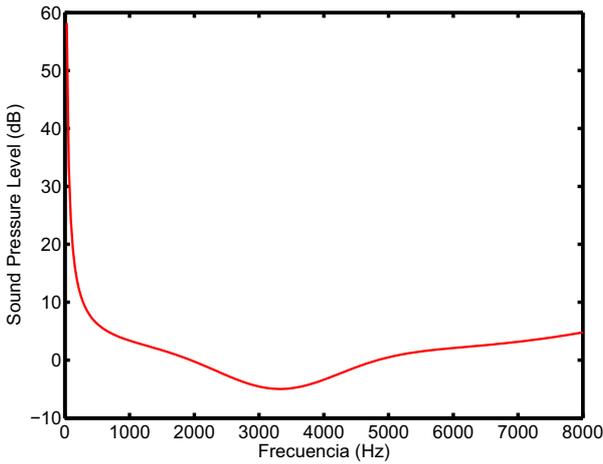


Figura 2. Umbral Absoluto de Audición correspondiente a una persona joven y saludable en un ambiente sin ruido.

de una persona con algún tipo específico de deficiencia auditiva, lo cual se verá mas en detalle en la Sección 3.

2.2 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 (Zwicker *et al.*, 1957). Esta clasificación surge del comportamiento mecánico del canal auditivo, que posee frecuencias características en las cuales presenta bajo rechazo, y alrededor de las cuales este rechazo aumenta rápidamente. Este comportamiento permite modelar al oído como un conjunto de filtros pasabanda, cada uno de ellos con una frecuencia central y ancho de banda independientes. Otra propiedad que caracteriza a estos filtros es que su ancho de banda aumenta junto a la frecuencia central del mismo, y es por ello que a bajas frecuencias un oído posee su mayor selectividad.

El enmascaramiento (Fletcher, 1940) es un fenómeno que se desprende del modelo de bandas críticas y ocurre cuando una componente de una determinada potencia impide que otra componente dentro de la misma banda crítica o una adyacente sea percibida. Existen dos tipos de enmascaramiento: el enmascaramiento simultáneo o espectral, y el enmascaramiento no simultáneo o temporal. Estos fenómenos se detallan a continuación:

**Enmascaramiento Espectral o simultáneo** Este fenómeno 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 al oído. Además, debe tenerse en cuenta que este es un fenómeno que no es espectralmente simétrico. En la Fig.3 se aprecian las características que influyen en este fenómeno. En la misma se aprecian en trazo verde fino y grueso, los umbrales de percepción en reposo y el debido a la presencia de componentes respectivamente. La primera de estas corresponde al nivel de energía mínimo para que una componente de una determinada frecuencia sea detectada, es decir el nivel por debajo del cual una componente *nunca es percibida*. La segunda corresponde al mismo umbral de percepción pero debido a la presencia de componentes de una determinada energía, es decir,

el umbral por debajo el cual una componente de menor energía es enmascarada por otra de mayor energía, llamada componente enmascaradora. En la Fig.3, la componente enmascaradora se representa utilizando trazo de color rojo, mientras que una componente enmascarada se representa en trazo color azul. Otro aspecto que se observa en esta Figura es que el umbral de enmascaramiento no es simétrico, ya que posee pendientes de  $+25\text{dB/Bark}$  y  $-10\text{dB/Bark}$ , por debajo y encima de la frecuencia del tono enmascarador respectivamente.

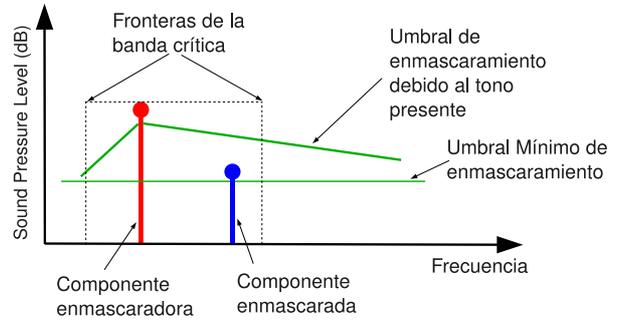


Figura 3. Parámetros característicos del enmascaramiento espectral.

Una representación de la distribución de las bandas críticas descriptas junto con su ancho de banda es presentado en la Fig.4.

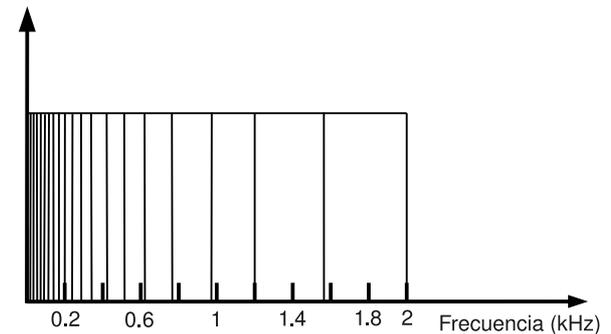


Figura 4. Ubicación y ancho de cada una de las bandas críticas.

Es posible distinguir entre cuatro tipos de enmascaramiento espectral: un tono siendo enmascarado por ruido (*Noise Masking Tone NMT*), ruido siendo enmascarado por un tono (*Tone Masking Noise TMN*), un tono siendo enmascarado por otro tono (*Tone Masking Tone TMT*), y por último, ruido enmascarado por ruido (*Noise Masking Noise NMN*).

El enmascaramiento espectral se utiliza en los sistemas perceptuales de compresión de datos, ya que las componentes que quedan enmascaradas por una de mayor intensidad son descartadas. De esta manera no son tenidas en cuenta para codificar un sonido. Si este fenómeno se varía de la forma adecuada, se puede simular el proceso de pérdida de selectividad que sufren las personas con algunos tipos de discapacidad auditiva. A modo de ejemplo, en la Fig. 5 se muestra en trazo continuo rojo la curva ATH en reposo, un conjunto de componentes enmascaradoras tonales (representadas mediante círculos de color magenta) y un conjunto de componentes enmascaradoras no tonales (representadas mediante círculos en color azul), las cuales generan el umbral de percepción que se representa en trazo continuo de color verde. El gráfico de la Fig. 5 representa

el umbral de enmascaramiento resultante de la aparición de un conjunto de componentes tonales y no tonales, que enmascaran a las que tienen energía menor al nivel de la curva en color verde.

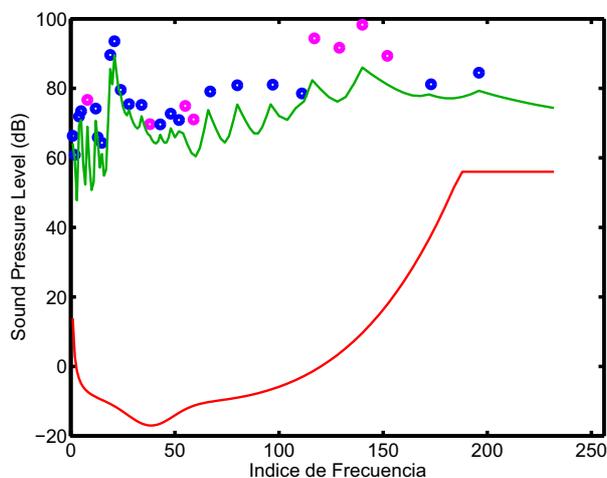


Figura 5. Detalle del Enmascaramiento Espectral.

*Enmascaramiento Temporal o no simultáneo* Este segundo tipo de enmascaramiento ocurre en el dominio del tiempo, y se debe a que el oído necesita un tiempo de recuperación o relajación 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 fenómeno se debe a la elasticidad que poseen las componentes del sistema auditivo humano, las cuales para un oído envejecido requieren un tiempo de reestablecimiento mayor. A su vez, este tiempo se divide en *preenmascaramiento temporal* y *postenmascaramiento temporal*. El primero de ellos se debe a la forma en la cual el cerebro adquiere información proveniente del canal auditivo, ya que si el mismo percibe dos sonidos, primero uno de baja intensidad y luego otro de mayor amplitud, separados por menos de  $10ms$ , la componente de mayor intensidad enmascara a la de menor amplitud. Esto se debe a que el cerebro prioriza la componente de mayor amplitud y por ello se enmascara la primera componente. Durante la persistencia del tono, cualquier componente de menor intensidad es enmascarada debido al fenómeno de enmascaramiento espectral detallado previamente. Pero una vez desaparecido este tono, el tiempo de relajación del oído es una función exponencial decreciente que relaciona el umbral de enmascaramiento y el tiempo, el cual puede ser de hasta  $200ms$ , para el oído de una persona joven en condiciones saludables. La Fig.6 detalla los tiempos asociados a los fenómenos previamente presentados.

En la Fig.6 se aprecian las variaciones del umbral de audición en función del tiempo, esto sucede porque en  $t = 0ms$  aparece un tono de  $SPL = 63dBV$ , como consecuencia de este tono se genera un *preenmascaramiento temporal* desde  $t = 0ms$  hasta  $t = 10ms$ . Durante la persistencia del tono, desde  $t = 10ms$  hasta  $t = 250ms$  existe enmascaramiento espectral para las componentes de menor energía. Por último, una vez desaparecido el tono, se presenta el *postenmascaramiento temporal*, el cual perdura hasta  $t = 410ms$ .

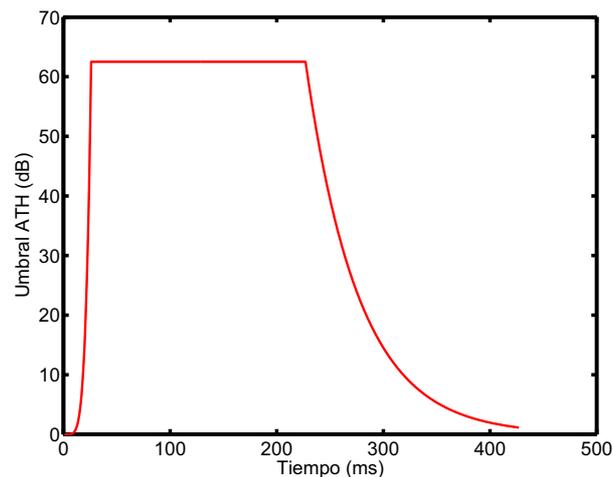


Figura 6. Detalle del enmascaramiento temporal.

### 3. DEFICIENCIAS A ESTUDIAR EN EL PRESENTE TRABAJO

Dado que este trabajo pretende estudiar como inciden algunos tipos de discapacidades auditivas sobre la inteligibilidad de los sonidos percibidos, es necesario introducir a las patologías que serán utilizadas en los experimentos. A lo largo de las siguientes Subsecciones se presentan las características más relevantes de cada patología, junto a la forma en que cada una de ellas influye en la percepción de los sonidos en una persona que la padece.

#### 3.1 Oído en condiciones ideales de salud

Con el fin de fijar una referencia para los experimentos a realizar es necesario implementar 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 simulación. Esto se debe a que el sistema en estas condiciones no afecta el sonido procesado con ningún tipo de deficiencia auditiva. La curva correspondiente a su Umbral Absoluto de Audición (ATH) se aprecia en la Fig.2.

#### 3.2 Oído afectado por una sordera severa

Este tipo de discapacidad (Vilchur, 1973), (Baer *et al.*, 2002) y (Timms, 2003) 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 del oído, el cual se modela sumando un rechazo de al menos  $50dB$  en las frecuencias superiores a la curva ATH, estos dos parámetros son función del tipo de discapacidad a simular y provocan que los sonidos que tengan un nivel por debajo de dicha curva no sean percibidos por el oyente. La Fig.7 (en trazo continuo de color azul) muestra un modelo de esta dolencia, donde se simula una discapacidad con una frecuencia de corte de  $1000Hz$ , mientras que en trazo rojo rayado se muestra la curva ATH de referencia, es decir, la curva correspondiente al oído de una persona joven, en un ambiente sin ruido.

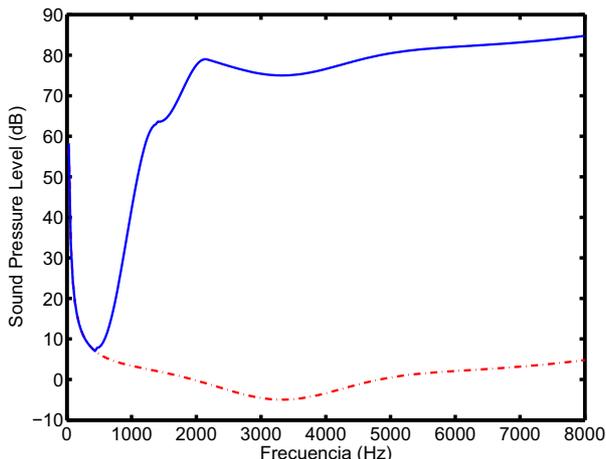


Figura 7. Curva SPL en función de la frecuencia para una discapacidad auditiva severa y de referencia.

3.3 Oído afectado por una discapacidad auditiva sensorineural

Para este tipo de discapacidad (Kim *et al.*, 1997), 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 ATH para este tipo de dolencia se representa en trazo azul continuo en la Fig.8. Nuevamente en color rojo, a modo de referencia, se presenta la curva ATH de un oído en buenas condiciones de salud.

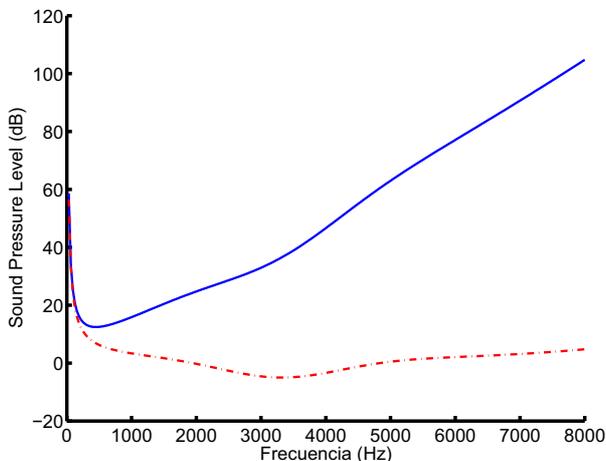


Figura 8. Funciones SPL en función de la frecuencia para una discapacidad sensorineural y de referencia.

3.4 Oído afectado por una discapacidad de origen traumático

Esta dolencia (Baer *et al.*, 2002), (Timms, 2003) y (Calupper and Fastl, 2000) 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 50dB entre  $f = 1500Hz$  y  $f = 3000Hz$ ). En la Fig.9, en trazo azul, se presenta la curva del ATH correspondiente a este tipo de dolencia junto con la curva ATH utilizada como referencia.

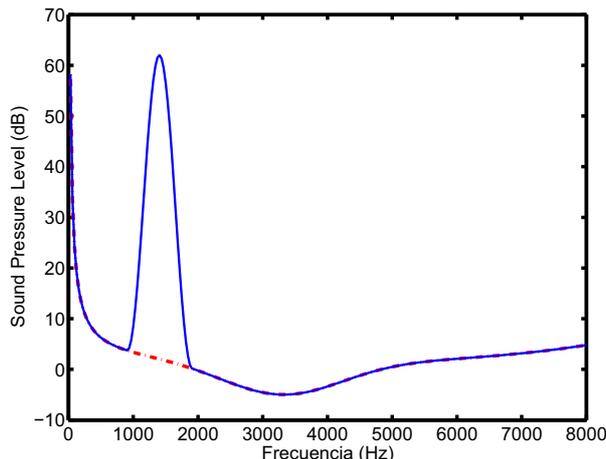


Figura 9. Funciones SPL en función de la frecuencia para una discapacidad de origen traumático y de referencia.

4. 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 3. Con este propósito, se genera un sistema que se divide en dos etapas. La primera toma la señal de audio  $x(t)$ , la procesa con el modelo perceptual presentado en la Sección 2, y como resultado se obtiene la información que debe ser utilizada para simular el comportamiento del oído, descartando la información que debido al enmascaramiento espectral no sea percibida por el modelo 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 3, a la salida de este bloque se obtiene como resultado la variación del umbral de enmascaramiento espectral asociada a la discapacidad bajo estudio, lo que permite conocer que información es descartada y cual debe ser tenida en cuenta a la hora de resintetizar el audio. Dado que, como se demuestra en Painter and Spanias (2000), el enmascaramiento espectral tiene mayor influencia en la pérdida de percepción de componentes que el enmascaramiento temporal, en esta instancia del trabajo sólo se modela el primer efecto.

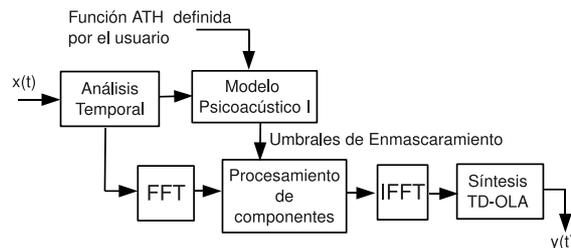


Figura 10. Diagrama en bloques del sistema.

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) con un solapamiento del 50%, y aplicarles la Transformada Rápida de Fourier (FFT, por sus siglas en inglés). En esta instancia, teniendo en cuenta la información proveniente del modelo auditivo, se realiza la selección de la información espectral que es percibida por el oído de la persona, en cada caso particular bajo estudio. De este modo, es eliminada toda

la información que el modelo auditivo considera que no es percibida. Por último, se realiza una Transformada de Fourier Inversa (IFFT) con el fin de obtener la señal procesada en el dominio del tiempo. El último paso del modelo consiste en un bloque que sintetiza la señal de voz procesada  $y(t)$ , mediante un algoritmo de suma solapada en el dominio del tiempo (*Time Domain-Overlap and Add* TD-OLA). Este modelo se muestra en la Fig. 10.

El modelo propuesto en este trabajo sólo implementa el efecto producido por el enmascaramiento espectral en el umbral de audición ATH. El efecto generado por el enmascaramiento temporal o no simultáneo no es implementado en el mismo. Por lo tanto, los tipos de discapacidad a simular por el mismo son los presentados en la Sección 3, ya que los mismos son casi exclusivamente causados por el enmascaramiento espectral.

#### 4.1 Mejoras al modelo

El sistema desarrollado en la Sección 4 presenta problemas, debido a que la señal  $x(t)$  es procesada sin considerar la naturaleza de la misma. Por ello, en esta subsección se describen las mejoras realizadas al sistema con el fin de optimizar el funcionamiento del mismo. A la hora de implementar el modelo de la sección anterior se hizo una suposición de que la señal  $x(t)$  bajo estudio era periódica, lo cual en principio es válido sólo en la zona central del fonema donde se asume la estacionariedad de la señal, pero en las fronteras entre fonemas esto no se cumple. Esto puede resolverse utilizando segmentos mas pequeños, o considerando un solapamiento mayor entre los segmentos. A continuación se detallan los problemas de implementación que debe afrontar el modelo y las formas de resolverlos.

**Dispersión Espectral y Efecto de Estacas** La **dispersión espectral** (*Leakage*) (Proakis and Manolakis, 1996) y (Huang *et al.*, 2001) es un efecto que aparece cuando una señal no periódica es analizada como si fuera periódica al momento de aplicar el algoritmo de FFT. Este efecto origina una dispersión espectral debida a que en la ventana de análisis no hay una cantidad entera de períodos de la señal.

En el caso de las señales de voz, el efecto de estacas (*picket fence effect*) (Proakis and Manolakis, 1996) y (Huang *et al.*, 2001), es una condición que aparece cuando la frecuencia de una componente que está siendo analizada no es un múltiplo entero de la frecuencia de muestreo ( $f_s$ ) dividida por la cantidad de puntos de la FFT ( $N$ ), es decir, si la frecuencia de la señal no es un múltiplo de la resolución de la FFT ( $f_s/N$ ). En este caso, el espectro de la señal obtenido por la FFT está afectado por los errores de amplitud  $e_A$  y de frecuencia  $e_F$ , que se presentan en (2) y (3).

$$e_f = f_{IN} - \frac{f_s \cdot k}{N} \quad (2)$$

$$e_A = \frac{N \sin[\pi(k-m)]}{2 \pi(k-m)} \quad (3)$$

Donde  $f_{IN}$  es la frecuencia de la componente de entrada,  $f_s$  es la frecuencia de muestreo,  $N$  las divisiones de la FFT y  $k$  el índice de la componente de la FFT mas cercana a la frecuencia de la señal de entrada. Por lo tanto, la componente  $k$ -ésima es la que minimiza los errores de (2) y (3). Dicha componente es la que suele utilizarse en el análisis de la señal.

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 Fig.11, donde  $F_{IN}$  (representada en línea de color azul) 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 líneas de colores verde y rojo. 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  $k$ -ésima, la mas próxima al tono afectado, la cual en la Fig.11 se representa mediante una línea en color rojo.

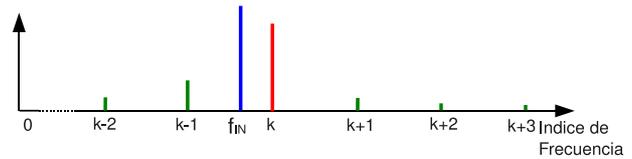


Figura 11. Error cometido por el efecto de estacas.

Este error es reducido considerando para el análisis no sólo la componente  $k$ -ésima de la FFT, sino que también se utilizan las componentes  $k-1$  y  $k+1$ , lo cual se aprecia en la Fig. 12, donde dichas componentes están graficadas en color rojo.

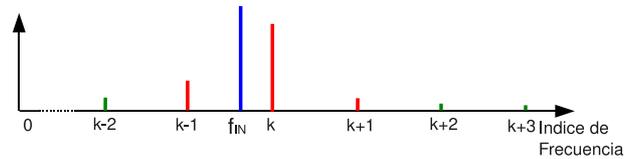


Figura 12. Corrección Implementada.

Cabe destacar que esta corrección sólo contempla una reducción en el error producido por 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. Este problema puede reducirse de tres formas:

1. **Aumentando el solapamiento entre segmentos:** Esto se realiza reduciendo la distancia entre segmentos consecutivos, desde 256 muestras ( $16ms$ ) hasta 128 muestras ( $8ms$ ), 64 muestras ( $4ms$ ) y 32 muestras ( $2ms$ ). Un enfo-

que similar es realizado en Timms (2003), donde el solapamiento entre segmentos puede ser de hasta un 99%. Al realizar esto la calidad de la voz sintetizada aumenta considerablemente y los artefactos desaparecen. La principal desventaja de esta solución es que la carga computacional aumenta proporcionalmente a la cantidad de segmentos a procesar, la que aumenta con el solapamiento entre las mismas.

2. **Considerando Componentes de Segmentos Consecutivos:** Con el fin de reducir las discontinuidades entre segmentos consecutivos, se realiza un promediado de la amplitud de las componentes de dicho segmentos. Las discontinuidades se reducen menos que en el método anterior, mientras que la carga computacional del sistema no aumenta por esta modificación.
3. **Combinación de 1 y 2:** Con el fin de optimizar la relación de compromiso entre reducción de discontinuidades y costo computacional, se propone una tercera solución que combina estos dos métodos. Esta solución consiste en promediar sólo las componentes que existan en un segmento y desaparezcan en el siguiente o viceversa, y además se aumenta el solapamiento entre los segmentos de modo tal que la distancia entre los mismos se reduzca de  $16ms$  a  $4ms$ . De esta forma se reducen las discontinuidades casi por completo, mientras que la carga computacional no aumenta de una manera significativa.

Un diagrama en bloques del sistema completo se aprecia en la Fig.13, donde el bloque llamado *sistema articulatorio* representa las consideraciones realizadas para reducir los errores por efecto de estacas y por las variaciones temporales de la señal procesada.

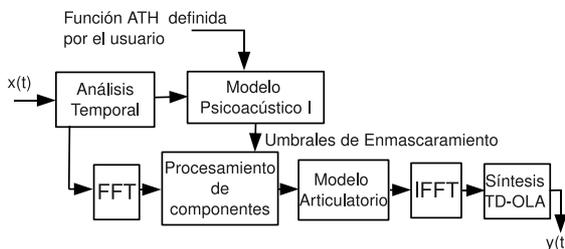


Figura 13. Diagrama en bloques del sistema con el corrector incorporado.

## 5. CONDICIONES EXPERIMENTALES

### 5.1 Base de datos utilizada

Con el fin de analizar el funcionamiento del sistema se realizaron experimentos con una base de datos de 100 logotomas (Miñana, 1969), pronunciadas por dos locutores: uno masculino y uno femenino. Los monosílabos utilizados son los propuestos por Miñana (1969), ya que como se demuestra en (Sommerhoff and Rosas, 2007), se caracterizan por contener la información necesaria para estudiar la inteligibilidad de la voz en idioma español. Esto se debe a que, como se muestra en (Quilis, 1999), la distribución de combinaciones que se aprecia en el siguiente listado se ajusta a la distribución existente en el idioma español.

- Consonante - Vocal (CV): 55 % de la base de datos.
- Consonante - Vocal - Consonante (CVC): 20 % de la base de datos.

- Vocal - Consonante (VC): 11 % de la base de datos.
- Consonante - Semiconsonante - Vocal (CScV): 5 % de la base de datos.
- Consonante - Semivocal - Vocal (CSvV): 1 % de la base de datos.

Las logotomas que componen la base de datos son presentadas en la Tabla 1.

Tabla 1. Lista de Logotomas publicada por Miñana.

res	va	la	quia	cen	sel	soir	cat	car	ted
die	dot	de	pec	be	sar	od	ce	net	sai
sap	co	sol	boc	nou	mou	ma	pac	at	al
ta	map	quo	ya	lor	diar	zat	yot	con	roi
cop	jol	me	leu	seir	ba	ven	dai	sia	nie
dior	to	cies	nos	no	pol	nal	sen	que	ep
fem	pe	loi	ras	fam	cat	tau	eir	es	les
rei	ser	nai	quei	get	teu	rios	cam	dep	eu
son	dog	rel	ram	au	la	nas	lied	rias	ye
lau	dac	ga	pa	tai	sau	der	seu	tel	ren

Estos permiten realizar estudios de inteligibilidad capaces de abarcar un gran porcentaje de combinaciones de fonemas utilizados en el idioma español.

### 5.2 Condiciones de grabación

Los datos fueron grabados utilizando un micrófono unidireccional con una sensibilidad de  $-52dBV$ , muestreados a 16 bits con una frecuencia de muestreo de  $16KHz$ . La sala de grabación está aislada acústicamente, obteniéndose una relación señal a ruido en los audios a procesar de al menos  $47dB$ . Los sonidos fueron articulados en español de Argentina. De esta forma, las grafías  $v$  y  $b$  se articularon como bilabiales (b), las grafías  $s$ ,  $z$  y  $c+e$  como alveolares (s). Con el fin de evitar efectos indeseados, las logotomas fueron grabadas con una separación de al menos  $2s$  entre ellas.

## 6. EXPERIMENTOS DE SIMULACIÓN DE DEFICIENCIAS

### 6.1 Casos de Estudio

Se realizaron tres experimentos subjetivos *Mean Opinion Score* (MOS) (ITU-T, 2003): el primero para analizar la calidad de los audios sintetizados, el segundo para estudiar la inteligibilidad y el tercero para analizar que sonidos son más afectados por determinados tipos de dolencias. Por otro lado, se realizó un experimento objetivo utilizando espectrogramas, con el propósito de estudiar si los espectros de las señales resintetizadas se ajustan a lo esperado para cada tipo y grado de dolencia. Todos los experimentos fueron realizados utilizando el sistema propuesto en la Sección 4, el cual corrige los errores producidos por el efecto de estacas y los causados debidos a la suposición de estacionariedad de las señales de voz estudiadas. La base de datos utilizada es la que se presenta en la Sección 5.

**Test Objetivo: Análisis utilizando espectrogramas** El espectrograma es una representación gráfica que permite analizar el comportamiento de las componentes espectrales a lo largo del tiempo. Este experimento se realiza con el fin de verificar que el simulador afecte cada una de las componentes según la condición específica que se esté simulando. Los espectrogramas son realizados utilizando la herramienta GNU *Praat*. El primer experimento analiza el funcionamiento del simulador

Figura 14. Espectrogramas correspondientes al logotoma *res* (en orden descendente): sin procesar por el simulador y procesada simulando un oído en buen estado de salud.

en condiciones ideales, es decir la simulación de un oído en buenas condiciones de salud. Como puede verse en la Fig. 14, donde se observan los espectrogramas de la logotoma *res* sin ser procesada por el simulador (figura superior), y procesada por el simulador utilizando una curva ATH en buenas condiciones de salud auditiva (figura inferior), se aprecia que los espectrogramas son similares, tal como era de esperarse. Para el caso de una sordera severa, en la Fig. 15 se presentan espectrogramas correspondientes al logotoma *car* en tres condiciones: en buenas condiciones de salud auditiva (figura superior), con una sordera severa simulada con un rechazo adicional a la curva ATH de  $50dB$  en frecuencias superiores a  $f = 1500Hz$  (figura central) y por último, con una sordera profunda simulada con un rechazo adicional a la curva ATH de  $50dB$  en frecuencias superiores a  $f = 750Hz$  (figura inferior). Los espectrogramas representados en la Fig. 15 muestran que el espectro de las señales resintetizadas se ajustan perfectamente a lo esperado para cada una de las discapacidades simuladas.

Por último, se estudia el caso de las discapacidades traumáticas, con este propósito, en las Fig. 16 y 17 se analizan dos logotomas. Las figuras muestran los casos de las logotomas *cel* y *cat* en dos condiciones simuladas: oído en buen estado de salud (en la parte superior de cada figura) y para un oído con una discapacidad traumática que le impide la percepción de sonidos entre  $f = 1000Hz$  y  $f = 2000Hz$  (en la parte inferior de cada figura). Se verifica que los sonidos sintetizados han sufrido una pérdida de información en los rangos de frecuencias afectados por la dolencia.

Si bien este experimento muestra que el sistema simula las discapacidades de una manera satisfactoria, se realizaron dos experimentos subjetivos con el fin de que evaluadores humanos analicen factores que no pueden ser apreciados mediante el uso de espectrogramas, como por ejemplo, la calidad de la voz sintetizada por el sistema y cómo cada una de las discapacidades simuladas afectan la inteligibilidad de los audios procesados. Por ello, 14 voluntarios (con su consentimiento informado) evaluaron una serie de audios tanto en **calidad** como también

Figura 15. Espectrogramas correspondientes al logotoma *car* (en orden descendente): simulando un oído saludable, simulando un oído con una discapacidad severa y simulando un oído discapacidad profunda.

Figura 16. Espectrogramas correspondientes al logotoma *cel* (en orden descendente): simulando un oído saludable y simulando un oído con una discapacidad traumática.

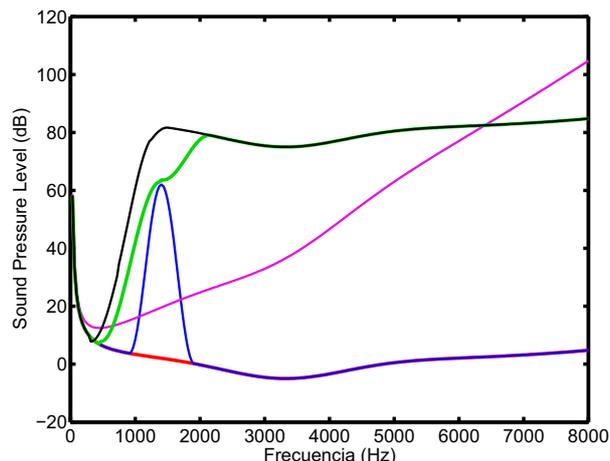


Figura 18. Curvas ATH simuladas en los experimentos.

Figura 17. Espectrogramas correspondientes al logotoma *cat* (en orden descendente): simulando un oído saludable y simulando un oído con una discapacidad traumática.

en **inteligibilidad**, calificando a cada uno 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. Además, los evaluadores debieron transcribir fonéticamente la interpretación realizada de cada uno de los monosílabos. De esta forma es posible analizar qué parte de la información que contienen los monosílabos se degrada debido a cada deficiencia simulada. El experimento estuvo compuesto por un conjunto de audios en las siguientes condiciones simuladas:

- **Audios sin ningún tipo de dolencia.** Para la simulación es utilizada una curva ATH como la que se muestra en trazo rojo en la Fig.18.
- **Audios afectados por una discapacidad auditiva severa.** Se utiliza la curva ATH presentada en color verde en la Fig.18.
- **Audios afectados por una discapacidad auditiva profunda.** Se utiliza la curva ATH presentada en color negro en la Fig.18.
- **Audios afectados por una discapacidad sensorineural.** Se utiliza la curva ATH graficada en color magenta en la Fig.18.
- **Audios afectados por una dolencia de tipo traumático.** Se utiliza la curva ATH graficada en color azul en la Fig.18.

**Test Subjetivo I: 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, dado que si en esta instancia se obtienen resultados elevados (en particular el caso del oído simulado sin ninguna discapacidad) significa que el algoritmo de procesamiento no afecta la señal y por lo tanto los únicos efectos que sufren los audios son debidos a la discapacidad simulada y a las condiciones de salud auditivas de los evaluadores (que se suponen en buenas condiciones). Los resultados se presentan en la Tabla 2, la cual presenta el valor medio de los datos ge-

nerados en el experimento, discriminados por sexo del locutor, y en conjunto. 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. El número de la primera columna de la Tabla 2 hace referencia a las siguientes condiciones:

- **1- Referencia.**
- **2- Oído en buen estado de salud.**
- **3- Trauma acústico bilateral,** considerando un rechazo adicional a la curva ATH de al menos  $50dB$  entre  $f = 1000Hz$  y  $f = 2000Hz$ .
- **4- Discapacidad sensorineural.**
- **5- Discapacidad severa,** considerando un rechazo adicional a la curva ATH de al menos  $50dB$  a partir de  $f = 1000Hz$ .
- **6- Discapacidad profunda,** considerando un rechazo adicional a la curva ATH de al menos  $50dB$  a partir de  $f = 700Hz$ .

Tabla 2. Test MOS de Calidad para la discapacidades simuladas.

	Loc. Masculino	Loc. Femenino	Conjunto
1	4.03	3.92	3.98
2	3.21	3.42	3.32
3	2.42	2.39	2.41
4	3.03	3.00	3.02
5	2.92	2.64	2.78
6	2.67	2.57	2.52

La Tabla 2 muestra que la calidad de los audios que se obtienen simulando el funcionamiento de un oído en buenas condiciones de salud es muy similar a la de los audios de referencia. Por otro lado, se aprecia que en general no se observa un deterioro importante en la calidad de los audios.

**Test Subjetivo II: Inteligibilidad** Este experimento analiza los problemas en la inteligibilidad de cada uno de los audios para cada una de las condiciones de discapacidad simulada detalladas en el experimento anterior.

**Tabla 3. Test MOS de Inteligibilidad para la discapacidades simuladas.**

	Loc. Masculino	Loc. Femenino	Conjunto
1	4.50	4.39	4.45
2	3.89	3.82	3.86
3	3.35	2.85	3.10
4	3.28	3.50	3.39
5	3.10	2.82	2.96
6	2.57	2.50	2.54

Los resultados son mostrados en la Tabla 3, donde se presenta el valor medio de las mediciones realizadas. Allí se puede apreciar que los audios que obtienen los resultados mas elevados (luego de las referencias), 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 acorde al hecho de que en ese tipo de discapacidades sólo se pierden algunas componentes de cada fonema, lo que genera una pérdida de la inteligibilidad que es apreciable sólo en algunos de ellos. Por ejemplo, en fonemas como: *s*, *z*, *f*, *m*, *n* y *l*, entre otros. En contraparte, fonemas que tienen una probabilidad de aparición elevada, como las vocales, no sufren alteraciones apreciables en esta discapacidad traumática simulada. De esta forma, el valor medio de la inteligibilidad medida por los evaluadores tiene un valor elevado.

El caso de una dolencia del tipo sensorineural 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 sólo los fonemas que tienen componentes relevantes en altas frecuencias se ven afectados. Esto se refleja en una incapacidad por parte de los evaluadores de percibir algunas componentes.

En los casos donde se analizan discapacidades severas se puede apreciar que se obtienen los resultados mas bajos, lo que se debe a la eliminación de componentes que tienen importancia a la hora de identificar un fonema. Tal como era de esperarse, la deficiencia simulada mediante un rechazo adicional a la curva ATH en frecuencias superiores a  $f = 1000Hz$  (fila 5 de la Tabla 3), obtuvo resultados superiores al caso de la deficiencia de carácter profunda que se presenta en la fila 6 de la Tabla 3. Esto se debe a que existen componentes en el rango de frecuencias entre  $f = 700Hz$  y  $f = 1000Hz$  que son indispensables a la hora de identificar un alto porcentaje de fonemas.

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 es afectada por la deficiencia auditiva es menor para el caso en que se escucha un fonema pronunciado por un hombre que el mismo fonema pronunciado por un niño o una mujer. Debido a que, por las dimensiones de su tracto vocal tienen frecuencias de resonancia mayores.

Los resultados obtenidos para otras discapacidades simuladas se ajustaron a lo esperado, ya que por ejemplo para sorderas mas severas que las estudiadas en las Tablas 2 y 3, como por ejemplo para dos deficiencias auditivas profundas simuladas mediante rechazos adicionales a la curva ATH a partir de  $f = 500Hz$  y  $f = 350Hz$ , respectivamente, en las que se

observó una reducción en la inteligibilidad acorde al grado de las discapacidades simuladas. Para otros tipos de dolencias, como otros tipos de discapacidades de origen traumático, se percibió la eliminación de formantes (Huang *et al.*, 2001) ubicados en la región afectada por la dolencia.

**Test Subjetivo III: Transcripción de logotomas** Con el fin de analizar qué fonemas son afectados por cada tipo de dolencia se pidió a los voluntarios que transcribieran su interpretación de cada audio analizado. Los resultados se muestran en las Tabla 4.

**Tabla 4. Porcentaje de acierto en la interpretación de logotomas.**

	Loc. Masculino	Loc. Femenino	Conjunto
1	92 %	90 %	91 %
2	82 %	78 %	80 %
3	50 %	21 %	35.5 %
4	31 %	35 %	28 %
5	29 %	21 %	30 %
6	17 %	17 %	17 %

En las Tabla 4, se aprecia que tanto en el caso de los audios de referencia como en los generados simulando un oído en buenas condiciones de salud, la interpretación fue casi perfecta. En el caso en que fueron simuladas dolencias, se pudo apreciar que ciertos fonemas eran interpretados como otros de manera sistemática. De esta forma, en el caso de la deficiencia de tipo traumática simulada, se pudo apreciar que la consonante *n* era percibida como *m* o *g* y que la consonante *p* como *t*. Otro aspecto a tener en cuenta es que para esta discapacidad en particular, se aprecia un deterioro en la interpretación de los fonemas emitidos por el locutor femenino. Esto se refleja en los resultados ya que se aprecia que el porcentaje de acierto en la interpretación de los fonemas para el locutor masculino duplica a los del locutor femenino. Para una deficiencia sensorineural, la consonante *q* fue interpretada como *j* y la *d* como *t*. Para este tipo de dolencias, el porcentaje de acierto en la interpretación de los audios pronunciados por locutores masculinos es levemente menor a los del locutor femenino. En el caso de las sorderas severas, el problema de interpretación se incrementó notablemente. Así para el caso de la sordera severa simulada previamente, además de existir errores de interpretación en los casos presentados para las otras discapacidades, existieron errores de interpretación en vocales. Así, la vocal *i* fue interpretada como *u* y para el locutor femenino, la *a* fue percibida como *o* casi en la mitad de sus apariciones. Esto se refleja en el porcentaje de acierto en la interpretación, el cual baja para ambos locutores. Cabe destacar, que los problemas de interpretación que aparecen para el locutor femenino se vuelcan en la Tabla 4, donde se aprecia que este es menor al del locutor masculino. Por último, para el caso de la sordera profunda, se apreció una disminución en la interpretación de los fonemas, lo que se aprecia en la última fila del cuadro referenciado.

## 7. 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 y la naturalidad de los audios sintetizados.

Como se observa en los resultados de los experimentos, el sistema simula los problemas perceptuales que aparecen en los diferentes fonemas y logotomas al ser percibidos por una persona con una determinada dolencia. Por ello el sistema puede ser utilizado como una herramienta para diseñar dispositivos que puedan ser utilizados por personas con diversos tipos de dolencias auditivas.

Los trabajos futuros apuntarán a incorporar al simulador los efectos generados por el enmascaramiento temporal, el cual no es implementado en este trabajo. Esta mejora permitirá aproximar más al comportamiento real el funcionamiento del sistema. Otro objetivo al que se apunta en el futuro es mejorar la calidad de los audios sintetizados, con el fin de que las variaciones respecto a los audios de referencia se deban sólo a las discapacidades simuladas en cada caso. De este modo, se podrá utilizar esta herramienta para el desarrollo de dispositivos de asistencia, reduciendo al mínimo la cantidad de evaluadores.

#### AGRADECIMIENTOS

Queremos agradecer a los miembros del Laboratorio de Comunicaciones (LAC) de la Facultad de Ingeniería de la Universidad Nacional de Mar del Plata, quienes participan en el Proyecto **Desarrollos Tecnológicos destinados a mejorar la Calidad de Vida de Personas con Capacidades Diferentes**, ámbito en el cual se desarrolla este Trabajo.

#### REFERENCIAS

- Baer, T., B.C.J. Moore and K. Kluk (2002). 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* **112**(3), 1133–1144.
- Calupper, J. and H. Fastl (2000). Simulation of hearing impairment based on the fourier time transform. *Proceedings of the IEEE ICASSP* **1**, 857–860.
- Denk, F., P.D. Agüero, A.J. Uriz, J.C. Tulli and E.L. Gonzalez (2010). Sistema de compresión de voz portátil basado en un dspic. *Proceedings of AST 2010* **1**, 1626–1637.
- Fletcher, H. (1940). Auditory patterns. *Modern Physics* **12**(1), 47–65.
- Huang, X., Acero, A. and Hon, H.W, Eds.) (2001). *Spoken Language Processing. A Guide of Theory, Algorithm, and System Development*. Prentice-Hall. Upper Saddle River, New Jersey 07458, Estados Unidos de America.
- ISO (1994). *International Organization of Standardization ISO/IEC JTC1/SC29/WG111 MPEG IS13813(1994): Generic Coding of Moving Pictures and Associated Audio*. International Organization of Standardization. Ginebra, Suiza.
- ITU-T, International Telecommunication Union, Ed.) (2003). *Series P: Telephone Quality, Telephone Installations, Local Line Networks: Methods for objective and subjective assessment of quality*. International Telecommunication Union. Ginebra, Suiza.
- Kates, J.M. (2010). Understanding compression: Modeling the effects of dynamic-range compression in hearing aids. *International Journal of Audiology* **49**(6), 395–409.
- Kim, D.W., Y.C. Park, W.K. Kim, S.J. Park, W. Doh, S.W. Doh, S.W. Shin and D.H. Youn (1997). Simulation of hearing impairment with sensorineural hearing loss. *Proc. of the 19th International Conference IEEE/EMBS* **1**, 1986–1989.
- Maniwa, K. and A. Jongman (2008). Perception of clear fricatives by normal-hearing and simulated hearing-impaired listeners. *Journal of the Acoustical Society of America* **123**(2), 1114–1125.
- Miñana, P., Ed.) (1969). *Compendio práctico de acústica*. Labor. Barcelona, España.
- Muñoz, C.M. Aguilera, P.B. Nelson, J.C. Rutledge and A. Gago (1999). Frequency lowering processing for listeners with significant hearing loss. *Proceedings of ICECS '99. The 6th IEEE International Conference on Electronics, Circuits and Systems* **2**, 741–744.
- Paarmann, L.D. and M.K. Wynne (1994). Bimodal perception of spectrum compressed speech. *Journal of the Acoustic of America* **95**(5), 2975–2976.
- Painter, T. and A. Spanias (2000). Perceptual coding of digital audio. *Proceedings of the IEEE* **88**(4), 451–513.
- Proakis, J.G. and Manolakis, D.G., Eds.) (1996). *Digital Signal Processing: Principles, algorithms and applications*. Prentice-Hall. Upper Saddle River, New Jersey 07458, Estados Unidos de America.
- Quilis, A., Ed.) (1999). *Tratado de fonología y fonética españolas*. Gredos. Madrid, España.
- Simpson, A.P. (1991). Characterization and simulation of impaired hearing: Implications for hearing aid design. *Ear and Hearing* **12**(6), 154–161.
- Slaney, M. (1998). *Auditory toolbox*. Technical Report 010. Palo Alto, Estados Unidos de America.
- Sommerhoff, J. and C. Rosas (2007). Evaluación de la inteligibilidad del habla en español. *Estudios filológicos* **42**, 215–225.
- Timms, O. (2003). *Speech Processing strategies based on the sinusoidal speech model for the profoundly Hearing Impaired*. Ph.D. Thesis. Tech Wissenschaften ETH. Zurich, Suiza.
- Uriz, A.J., P.D. Agüero, J.C. Tulli, E.L. Gonzalez and F. Denk (2009). Desarrollo de un sistema de compresión de voz portátil para pacientes discapacitados. *Proc. of SABI 2009*.
- Vilchur, E. (1973). Signal processing to improve speech intelligibility in perceptive deafness. *Journal of the Acoustical Society of America* **53**(6), 1646–1657.
- Widex (2003). Manual técnico del dispositivo bravo.
- Widex (2004). Manual técnico del dispositivo senso plus sic.
- Zwicker, E., G. Flottorp and S.S. Stevens (1957). Critical band width in loudness summation. *JASA* **29**(3), 548–557.