

Desarrollo de un Sistema de Compresión de Voz portátil para Pacientes Discapacitados

Alejandro J. Uriz, Pablo D. Agüero, Juan C. Tulli, Esteban L. Gonzalez y Francisco Denk.
Universidad Nacional de Mar del Plata, Facultad de Ingeniería, ajuriz@fi.mdp.edu.ar

Resumen— Las pérdidas de audición son una dolencia que afecta a un gran porcentaje de la humanidad, y las soluciones que se encuentran en el mercado tienden a ser efectivas pero costosas. Este trabajo se enfoca en el desarrollo de un sistema capaz de simular problemas de audición en seres humanos, y en base a este análisis, desarrollar un dispositivo portátil capaz de ayudar al paciente usando un dsPIC producido por Microchip. Para ello se ha diseñado un modelo capaz de simular pérdidas de audición mediante un modelo de filtros, y con él se ha experimentado con dos tipos de compresión: lineal y SPINC. Los resultados subjetivos experimentales demuestran las ventajas de la utilización de la función SPINC sobre la compresión lineal.

Palabras clave— discapacidades auditivas, conversión de voz, compresión de voz, síntesis de voz.

I. INTRODUCCIÓN

Las pérdidas de audición son una dolencia que en Argentina aqueja, según datos del Censo del 2001, a 300.000 de personas.

Existen diversos tipos y grados del problema, pudiendo ser clasificados por su origen en congénito, debido a envejecimiento o a un accidente, entre otros. Sin embargo, una forma más útil de clasificarlas es a través del umbral de audición que presentan [1]-[3]. Un nivel de sordera grave tiene un umbral de audición de 90dB, mientras que en una sordera leve resulta de alrededor de 25dB. A su vez, debe tenerse en cuenta que generalmente un problema de audición sólo afecta a un rango de frecuencias dentro de la zona de audición, siendo el caso más común una pérdida que comience en las más altas frecuencias y vaya hacia las bajas frecuencias.

Para tratar con los distintos tipos de sordera se han desarrollado un conjunto de técnicas que son aplicadas a diversos dispositivos colocados en distintas partes del oído, tales como audífonos e implantes cocleares. Son ejemplos de estos dispositivos los ofrecidos por compañías como Widex y Samsung, con una gran variedad de productos, los cuales van desde audífonos analógicos (que pueden poseer sólo un circuito amplificador), hasta equipos digitales con tecnología de punta [4], [5]. Estos últimos permiten realizar tareas tales como compresión de voz, filtrado dinámico de ruido, opciones de funcionamiento configurables y acoplamiento para conversaciones telefónicas, entre otras. Otro aspecto que es considerado en estos diseños es la ergonomía, ya que estos productos están pensados para su uso continuo en el cuerpo del paciente. Por ello, las compañías mencionadas tienen en cuenta que el equipo no provoque reacciones alérgicas a la piel del paciente, que la forma del dispositivo no afecte al oído y que el dispositivo sea inmune a la humedad debida al sudor o la lluvia.

Una de las principales ventajas de los dispositivos

digitales es que permiten obtener soluciones a la medida del paciente. En especial, pueden ser usados para resolver problemas que estén relacionados con el envejecimiento del oído o con la pérdida de audición en un rango de frecuencias en particular. Este tipo de casos pueden ser asistidos utilizando técnicas de compresión de voz, las cuales permiten acondicionar la señal que ingresa al oído del oyente con el fin de llevar el rango de frecuencias donde exista una “zona muerta” del órgano hacia zonas donde la información pueda ser oída.

La compresión se lleva a cabo mediante una función que mapea de una manera adecuada las frecuencias de la zona afectada hacia la nueva región. De este modo, pueden encontrarse versiones lineales basadas en la Transformada Rápida de Fourier (FFT) [6]-[8]; versiones basadas en modelos sinusoidales [9], [10]; e incluso usando aproximaciones basadas en bancos de filtros [11], [12].

En este trabajo se apunta a desarrollar un sistema capaz de simular distintos grados de pérdidas auditivas y técnicas que logren mitigar dichos problemas. El objetivo final del trabajo consiste en desarrollar un dispositivo que sea capaz de integrar las técnicas utilizadas en un dispositivo DSP portátil, del tipo dsPIC. De este modo, se hace posible comparar el producto desarrollado con los existentes en el mercado, apuntando a mejorar las especificaciones y reducir su costo.

El artículo se organiza de la siguiente manera: en la Sección 2 se hace referencia a las dolencias que el sistema apunta a resolver. La Sección 3 presentan los métodos del estado de la cuestión a desarrollar y se implementa el algoritmo de compresión utilizado. A lo largo de la Sección 4 se realizan experimentos para validar los métodos implementados. Finalmente en la Sección 5 se hace un análisis de la factibilidad para implementar el algoritmo desarrollado en un microcontrolador dsPIC, con el fin de obtener un dispositivo portátil. Por último se desarrollan las conclusiones del trabajo.

II. DESCRIPCIÓN DEL ENFOQUE DE LAS PATOLOGÍAS

En esta sección se presenta un breve resumen de las dolencias que se van a estudiar a lo largo del trabajo. Este trabajo está enfocado en resolver dolencias similares a un envejecimiento del órgano auditivo, las cuales se caracterizan en un aumento en el umbral de audición y la aparición de “zonas muertas”, especialmente en las frecuencias altas y medias. Estas representaciones son desarrolladas en [3], [13]. Esto trae aparejado que algunos fonemas, en especial los fricativos [14], no puedan ser correctamente oídos, ya que una parte de su información espectral está siendo eliminada por la reducción de la audición del órgano. De los estudios realizados en [2] se

desprende la Fig. 1, que muestra la curva de audición de un oído en buen estado y otro con problemas graves.

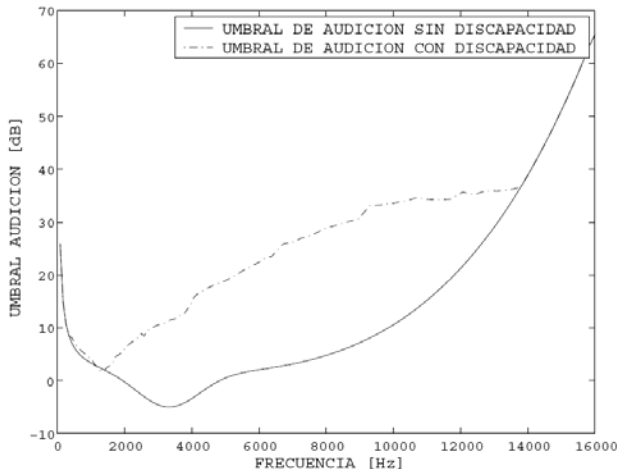


Fig. 1: Respuestas en frecuencia del oído de personas en diferentes condiciones de salud auditiva.

Como se mencionó en el párrafo anterior, los problemas auditivos se deben a que el oyente percibe menos formantes u otra información frecuencial que resulta necesaria para poder distinguir entre fonemas.

Estudios previos indican que solo son necesarios los primeros formantes para poder distinguir entre algunos fonemas sonoros, e incluso demuestran que la energía se concentra en un alto porcentaje para fonemas sonoros entre 0 y 1KHz para una persona con voz grave [6]. A su vez, para fonemas fricativos, dichos formantes se encuentran en frecuencias mas elevadas. Cabe destacar además que la información necesaria para reconocer al locutor se halla ubicada en el rango de frecuencias entre los 1000Hz y los 2000Hz [15].

En consecuencia resulta factible intentar trasladar los formantes hacia el rango de frecuencia donde existe un resto auditivo, para que el oyente perciba el mensaje de una manera correcta. Dicha tarea es realizada por el sistema de compresión de voz que se presenta en la siguiente sección.

III. Métodos implementados

Esta sección se divide en dos partes. En la primera se introduce un modelo que implementa herramientas capaces de simular las deficiencias auditivas planteadas en el capítulo previo. La segunda parte de la sección se enfoca en la presentación de un algoritmo que implementa un método de compresión de voz con el fin de mejorar la inteligibilidad de la señal que percibe el oyente.

A. Modelado de la discapacidad auditiva

Para comenzar el estudio es necesario analizar el problema de la pérdida auditiva. A través del estudio de esta deficiencia es posible construir un modelo del oído como un banco de filtros con diferentes umbrales de audición [13], [16], [17]. De esta forma, toda aquella señal con potencia inferior al umbral resulta eliminada del análisis. Esto nos permitirá obtener la señal que esta siendo oída por una persona con distintos grados de discapacidad.

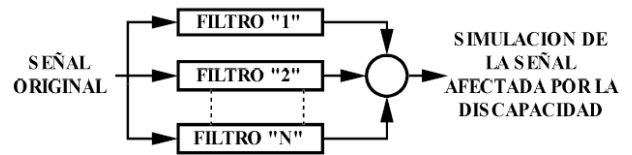


Fig. 2: Esquema del banco de filtros usado para la simulación de discapacidades auditivas.

La Fig. 2 presenta el sistema modelado. Para implementar el modelo, se hace uso de diferentes herramientas de modelado psicoacústico que se complementan con los estudios realizados por diversos científicos acerca de la forma del filtro de audición para personas con distintos tipos de sordera [18]. Los resultados de este desarrollo nos permiten conocer como oye una persona con una determinada discapacidad auditiva, y de esta manera se podrán construir herramientas más efectivas para mejorar la audición. Este modelo consiste en modelar las pérdidas auditivas como un conjunto de filtros pasabanda. Un ejemplo de esto, es modelar el oído mediante un filtro paso bajos, el cual atenúa las componentes de alta frecuencia que un oído envejecido no es capaz de percibir.

B. Algoritmo de compresión de voz

Los algoritmos de compresión espectral se basan en escalamientos y desplazamientos de frecuencia. Para cada una de las frecuencias de entrada se establece una correspondencia con una frecuencia de salida a través de un factor de escala y un desplazamiento en frecuencia. Tanto el factor como el desplazamiento pueden obedecer una ley lineal, logarítmica, u otra mas compleja con respecto a la frecuencia de entrada.

En la literatura se pueden encontrar diferentes funciones que relacionan la frecuencia de entrada con la frecuencia de salida [5], [6], [19]. En general, la compresión en las bajas frecuencias es mínima y la compresión en altas frecuencias obedece una ley cercana a la logarítmica.

En nuestro caso utilizaremos una función de mapeo frecuencial que responde al arco tangente, la función SPINC [19], mediante la cual las frecuencias de entrada (f) y de salida $\Phi(f)$ se relacionan según la Ec. 1.

$$\phi(f) = 1414 \arctg(f / 1414 \text{Hz}) \text{ Spinc. (1)}$$

Esta función presenta diversas ventajas en su comportamiento tanto en bajas como altas frecuencias.

Diversos estudios, entre los cuales se puede mencionar el realizado por Simpson [20], indican que los formantes de mas baja frecuencia, inferiores a 300 Hz, no presentan variaciones significativas entre locutores con respecto a lo que ocurre con los formantes de frecuencias superiores. En consecuencia, resulta deseable el comportamiento de la función SPINC, la cual preserva la ubicación de dichos formantes. Otra ventaja de esta función, es que es posible comprimir todo el espectro de voz en un ancho de banda 2200Hz. Mientras que si se utiliza compresión lineal con un factor de compresión 1,3, para transmitir 20KHz se necesita un ancho de banda de aproximadamente 15KHz. Pero esta ventaja en la compresión del ancho de banda utilizado trae aparejado una pérdida de selectividad para las altas frecuencias, ya que la función SPINC comprime por un

factor unitario entre 0 y 1414Hz y luego comprime el resto del espectro entre 1414Hz y 2,2KHz. Esto reduce la distorsión de los primeros formantes de la señal de voz y comprime por un factor mayor la zona con menor información acústica, es decir donde se hayan los formantes superiores.

En este trabajo se implementa el algoritmo de compresión utilizando el algoritmo propuesto por [7]. El mismo consiste en un análisis solapado de la señal usando la transformada rápida de Fourier (FFT). La compresión espectral se realiza trasladando los picos frecuenciales usando la función SPINC. Finalmente, se aplica un filtrado pasabajos y se realiza un espejado del espectro para recuperar la simetría hermitiana. Este último paso es importante para obtener una señal real a través de la transformada inversa de Fourier (IFFT), para la reconstrucción posterior usando el algoritmo de suma solapada.

En este caso se utilizan los siguientes parámetros:

- Filtrado de preénfasis para realzar las altas frecuencias.
- Transformación al dominio de la frecuencia usando una FFT de 512 puntos. De esta forma, se obtiene una resolución en frecuencia de 31,25Hz para una frecuencia de muestreo de 16KHz.
- Se utilizan segmentos de señal (frames) de 512 puntos con un solapamiento de 175 puntos, para el enventanado se usa una ventana de Hamming. No se hace distinción de zonas sonoras / sordas.
- Por último se realiza un filtrado pasabajos a diferentes frecuencias según la pérdida auditiva.

De esta forma se obtiene un sistema capaz de comprimir los sonidos de un determinado rango donde el oyente presente problemas hacia otro donde sea capaz de oír. Este sistema tiene una carga computacional que depende fundamentalmente de la función FFT implementada, es por ello que se busca un compromiso entre la resolución en frecuencia del sistema y la carga computacional.

IV. Experimentos Realizados

Con el fin de cuantificar la mejora introducida por el método, se realizaron experimentos para analizar la inteligibilidad y la calidad de los audios obtenidos.

Para ello, se utilizó el algoritmo presentado en la Sección 3.B para procesar los audios de 4 locutores: 2 mujeres y 2 hombres. Por otro lado se utiliza una serie de sonidos que el discapacitado debe escuchar en su vida cotidiana (bocinas de autos, advertencias ferroviarias, etc). El comportamiento del sistema se estudia en una condición de sordera simulada mediante un filtro pasabajos con una frecuencia de corte de 870Hz. De este modo, el sonido sin compresión carece de la información necesaria para que algunos fonemas sean inteligibles, como es el caso de los fonemas fricativos.

El experimento consta de un grupo de oyentes sin discapacidades auditivas que puntúan de 1 a 5 en inteligibilidad (5: perfectamente inteligible, 1: ininteligible) y calidad (5: calidad excelente, 1: calidad pésima, deteriorada por la manipulación de los datos) una serie de archivos de audio los cuales corresponden a los siguientes casos:

- Filtrado pasabajos, con reconstrucción usando compresión lineal y factores de compresión 1,3 y 1,7.
- Filtrado pasabajos, con reconstrucción usando compresión SPINC y factores de 1,3 y de 1,0.

A continuación se presenta la Tabla 1 con los resultados obtenidos:

	INTELIGIBILIDAD	CALIDAD
COMPRESION LINEAL 1,3	1,7	1,7
COMPRESION LINEAL 1,7	1,7	1,8
COMPRESION SPINC 1,3	2,1	1,9
COMPRESION SPINC 1,0	2,4	2,2

Tabla 1. Resultados del experimento.

Los resultados muestran una clara ventaja tanto en la inteligibilidad como en la calidad para el algoritmo de compresión SPINC por sobre el uso de compresión LINEAL. Además, el algoritmo SPINC con un factor de compresión unitario es el que presenta mejores resultados. Esto se debe a que los restantes algoritmos comprimen los formantes más bajos de la señal de voz y por ello resulta distorsionada. Esto hace que en algunos casos la señal que oye el locutor sea ininteligible. Del mismo modo, para sonidos de advertencia se observó que sin aplicar compresión era imposible percibirlos. El caso mas relevante fue el de una señal de advertencia de un paso a nivel, la cual se pudo escuchar perfectamente usando compresión SPINC, mientras que sin compresión o usando compresión lineal, la advertencia no pudo ser percibida.

V. Implementación en un Dispositivo DSP

Dado que las evaluaciones realizadas arrojaron resultados significativos, se apunta a implementar el algoritmo mediante el uso de un dispositivo dsPIC de Microchip [21], [22]. Para comenzar el diseño, se investigó acerca de los productos similares existentes en el mercado. Esto arroja como primera conclusión que hay dos grandes ramas de dispositivos de asistencia medica, los analógicos y los digitales. Los primeros se basan en sistemas de filtros y amplificadores que realizan un limitado número de funciones. En cambio, los dispositivos digitales trabajan con una gran cantidad de funciones debido al grado de sofisticación alcanzado por los circuitos integrados. De este modo es posible observar audífonos comerciales analógicos que se limiten sólo a amplificar una señal mientras que existen audífonos digitales como el BRAVO o el INTEO, ambos de la empresa WIDEX, que poseen funciones como control de ruido, compresión de voz y ecualización, entre otras. Otra gran ventaja de los dispositivos digitales es que permiten ajustar el equipo a las necesidades del oyente con solo reprogramarlo, evitando calibraciones extensas. Por lo comentado, estos dispositivos de asistencia son muy versátiles, y esto se ve reflejado en un gran aumento en el costo del dispositivo. Es por ello que los dispositivos de tecnología digital tienen precios superiores a los US\$1000. Por otro lado, debe tenerse en cuenta que estos dispositivos trabajan con dispositivos DSP, los cuales necesitan una cantidad de componentes externos. Por otra parte, el dispositivo dsPIC propuesto tiene una gran integración de componentes en su interior, reduciendo así el volumen del equipo. Esta característica hace viable la idea de construcción del dispositivo de asistencia en base a un dsPIC de Microchip.

En particular, el punto mas importante a tener en

cuenta a la hora de la selección del dispositivo es que combina la capacidad para el manejo de señales de un DSP con una arquitectura versátil como es la de los microcontroladores PIC. Entre otras características destacadas se encuentran:

- Gran cantidad de memoria de programa. Lo que es apropiado para el uso de compiladores cruzados.
- Gran velocidad de procesamiento (superior a 20 millones de instrucciones por segundo).
- Bajo costo; El dsPIC utilizado para la simulación ronda los U\$S 10. Esta es una gran ventaja en comparación con dispositivos DSP comerciales.
- Bus de datos de 16 bits.

Además debe tenerse en cuenta que la documentación acerca de los dispositivos Microchip y sus librerías están disponibles en Internet sin costo alguno.

Con el fin de realizar la implementación, la cual se observa en la Fig. 3, se seleccionó el dispositivo dsPIC30F2010 el cual se detalla en [21], [22], haciendo uso de las librerías matemáticas y de la librería para procesamiento de señales [23]-[25]. El algoritmo de compresión de voz implementado es el descrito en la Sección 3. Mediante la herramienta MPLAB se pudo simular el desempeño del sistema, comprobando que es posible desarrollar un dispositivo que trabaje en tiempo real. Este dispositivo es posible adaptarlo al oyente, con solo afectar el valor del factor de compresión y del desplazamiento a realizar. Otro factor a tener en cuenta es el costo, dado que en comparación con productos existentes en el mercado, el producto diseñado tendrá un costo mucho menor.



Fig. 3: Esquema del dispositivo propuesto.

VI. CONCLUSIONES

A lo largo de este trabajo se hizo un desarrollo mediante el cual se obtuvo el diseño de un dispositivo capaz de realizar compresión de voz para oyentes con deficiencias auditivas. Se pudo demostrar que mediante algoritmos más potentes como es el SPINC y dispositivos con mayor capacidad de cómputo es posible mejorar los dispositivos similares que se encuentran actualmente en el mercado. Esto se evidencia en los experimentos realizados ya que el algoritmo utilizado mejora el rendimiento de la compresión lineal, la cual es usada comercialmente. Otro aspecto relevante es que el desarrollo se realizó usando un dispositivo comercial dsPIC, perteneciente a una línea de dispositivos que está en constante evolución, lo cual permitirá seguir mejorando el equipo en paralelo con el desarrollo de estos dispositivos programables.

REFERENCIAS

[1] E. Villchur, "Signal processing to improve speech intelligibility in perceptive deafness", en *Journal of the Acoustical Society of America*, vol. 53(6), pp. 1646-1657, 1973.

[2] T. Baer, B. C. J. Moore y K. Kluk, "Effects of low pass filtering on the intelligibility of speech in noise for people with and without dead regions at high frequencies", en *Journal of the Acoustical Society of America*, vol. 112(3), pp. 1133-1144, 2002.

[3] E. Zwicker, G. Flottorp y S. S. Stevens, "Critical Band Width in Loudness Simulation", en *JASA*, vol. 29(3), pp. 548-557, 1957.

[4] Widex Inc., "Manual Técnico de dispositivo BRAVO B32", 2003.

[5] Widex Inc., "Manual Técnico del dispositivo SENSO PLUS SIC", 2004.

[6] L. D. Paarmann y M. K. Wynne, "Bimodal perception of spectrum compressed speech", en *Journal of the Acoustical Society of America*, vol. 95(5), pp. 2975-2976, 1994.

[7] O. Timms, "Speech Processing strategies based on the sinusoidal speech model for the profoundly Hearing Impaired", Ph.D. Thesis, 2003.

[8] K. Yasu, K. Kobayasi, K. Shinohara, M. Hishitani, T. Arai y Y. Murahara, "Frequency compression of critical band for digital hearing aids", en *Proc. of China-Japan Joint Conference on Acoustics*, pp. 159-162, 2002..

[9] T. F. Qualtiery y R. J. McAulay, "Speech Analysis/Synthesis based on a Sinusoidal Representation", en *IEEE Trans. On Speech and Audio Processing*, vol. 34, No.4, pp. 744-754, 1986.

[10] C. M. Aguilera Muñoz, P. B. Nelson, J. C. Rutledge y A. Gago, "Frequency Lowering processing for listeners with significant hearing loss", en *IEEE Trans. On Speech and Audio Processing*, vol. 34, No. 4, pp. 744-754, 1986.

[11] D. S. Chaudari y P.C. Pandey, "Critical band splitting of speech signal for reducing the effect of spectral masking in bilateral sensorineural hearing impairment", en *Proc. of the 5th International Symposium on Signal Processing and its Applications*, vol 1, pp. 119-122, 1998.

[12] K. Yasu, K. Kobayashi, K. Shinohara, M. Kohshi, T. Arai y Y. Murahara, "Critical-Band compression method for digital hearing aids", en *Journal of the Acoustic Society of Japan*, vol. 25, No. 1, pp.61-63, 2004.

[13] B. R. Glasberg y B. C. J. Moore, K. Shinohara, M. Kohshi, T. Arai y Y. Murahara, "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments", en *Journal of the Acoustical Society of America*, vol. 74(4), pp. 1120-1133, 1986.

[14] X. Huang, A. Acero y H.W. Hon, "Spoken Language Processing. A Guide of Theory, Algorithm, and System Development", 2001.

[15] O. Turk y L. M. Arslan, "Subband Based Voice Conversion", en *ISCLP 2002*, pp. 289-292, 2002.

[16] B. Hackstein y M. Schwartz, "Implementation and Visualisation of MP3-Algorithm in Matlab", 2006

[17] D.W. Kim, Y.C. Park, W.K Kim, S.J. Park, W. Doh, S.W. Shin y D.H. Youn, "Simulation of hearing impairment with sensorineural hearing loss", en *Proceedings of the 19th International Conference of the IEEE*, vol.5, pp.1986-1989, 1997.

[18] M. Slaney, "Auditory Toolbox. Version 2", en *Technical Report #1998-010. Interval research Corporation*, 1998.

[19] E. Terhardt, "The SPINC Function for scaling of frequency in Auditory Models", en *Journal of Acustica*, vol.77, pp.40-42, 1992.

[20] A. P. Simpson, "Dynamic consequences of differences in male and female vocal tract dimensions", en *Journal of the Acoustical Society of America*, vol. 109(5), pp. 2153-2164, 2001.

[21] Microchip Inc., "dsPIC30F Family overview. DsPIC High Performance, 16-bit Digital Signal Controllers". 2004.

[22] Microchip Inc., "dsPIC33FJ32GP302/304, dsPIC33FJ64GPX02/04, dsPIC33FJ128GPX02/04 Data Sheet. High Performance, 16-bit Digital Signal Controllers", 2009.

[23] Microchip Inc., "dsPIC30F Speech Recognition Library user's guide", 2004.

[24] Microchip Inc., "dsPIC30F Math Library", 2004.

[25] Microchip Inc., "dsPIC DSC Noise supsression Library user's guide", 2008.