

Audífonos binaurales individualizados de rango extendido optimizados para localización espacial e inteligibilidad del habla en entornos sonoros complejos

Individualized extended range binaural hearing aids optimized for spatial localization and speech intelligibility in complex acoustic environments

Presentación: 4 y 5 de Octubre de 2022

Doctorando:

Fermín Scaliti

Centro de Investigación y Transferencia en Acústica (CINTRA), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad Tecnológica Nacional - Facultad Regional Córdoba (UTN - FRC) - Argentina
fscaliti@frc.utn.edu.ar

Director:

Fabián C. Tommasini

Codirector:

Diego A. Evin

Resumen

La restauración de la audición es posible mediante el uso de audífonos, que son dispositivos de tecnología asistiva que amplifican los sonidos para compensar la disminución de la sensibilidad auditiva. Sin embargo, la calidad sonora que proporcionan actualmente es subóptima. Por otro lado, la hipoacusia degrada las habilidades de localización espacial de los sonidos y la conservación de todas las claves de audición espacial es uno de los principales problemas de los audífonos actuales. En este trabajo se presenta un plan de investigación y desarrollo que busca realizar aportes al estado del arte en aspectos aún no resueltos en los audífonos digitales actuales, como son, la mejora de la inteligibilidad del habla en ambientes sonoros complejos y la localización espacial de fuentes sonoras junto con la mejora de la naturalidad auditiva. Para ello se propone desarrollar un modelo de audición binaural para audífonos que incluya algoritmos de realce del habla y clasificación de escenas sonoras, e incorpore una función de corrección individualizada dependiente de la dirección de incidencia de los sonidos.

Palabras clave: Audífonos binaurales, audición espacial, inteligibilidad del habla.

Abstract

Hearing restoration might be possible by using hearing aids, which are assistive technology devices that amplify sounds to compensate for the decrease in hearing sensitivity. However, the sound quality that these devices deliver is still suboptimal. Apart from that, hearing loss can impair spatial hearing, and the conservation of all spatial cues is one major unsolved issue of

modern hearing aids. This work presents a research and development plan that aims to carry out a contribution to the state-of-the-art in issues yet to be solved in current hearing aids, like speech intelligibility enhancement under noisy environments and sound source localization along with hearing naturalness improvement. To this end, it is proposed to develop a binaural hearing model for hearing aids that includes algorithms for speech intelligibility enhancement and sound scene classification, and incorporates an individualized, direction-dependent spatial hearing correction function..

Keywords: Binaural hearing aids, spatial hearing, speech intelligibility

Introducción

La pérdida auditiva o hipoacusia impacta directamente en la comunicación y en la vida diaria (Ciorba et al., 2012), lo que la convierte en un serio problema de salud y también social. En la gran mayoría de los casos, la restauración de la audición es posible mediante el uso de audífonos, que son dispositivos de tecnología asistiva que amplifican los sonidos para compensar la disminución de la sensibilidad auditiva (Popelka et al., 2016). Sin embargo, la calidad sonora que proporcionan actualmente es subóptima (Denk et al., 2017).

La audición binaural –o audición con los dos oídos– es un proceso que permite identificar la dirección espacial de dónde provienen las señales sonoras, filtrar el ruido no deseado, y reconocer una voz entre muchas y/o con ruido de fondo (Werner et al., 2012). En este complejo proceso, la persona extrae diferentes indicios perceptuales de las ondas sonoras que llegan a los dos oídos. Cuando un sonido proviene desde uno de los laterales, las ondas recorren diferentes distancias antes de llegar a los tímpanos, dando lugar a una diferencia interaural de tiempo (interaural time difference, ITD). Además, uno de los oídos queda oculto por la cabeza y se produce una “sombra acústica” dando origen a una diferencia de nivel de presión sonora entre ambos oídos, denominada diferencia interaural de nivel (interaural level difference, ILD). Tanto la ITD como la ILD, son los principales indicios para localizar una fuente sonora en el plano horizontal o azimutal, y son llamadas claves binaurales. En tanto que, en el plano vertical y en el cono de confusión (Wallach, 1939), la persona identifica indicios espectrales producidos por el filtrado natural que realiza el cuerpo humano y, principalmente, el pabellón auricular, sobre el sonido según la dirección de incidencia. Estos indicios, llamados claves monoaurales, son importantes para determinar la ubicación de una fuente sonora en el plano vertical (Asano et al., 1990). Las claves binaurales y monoaurales en conjunto están presentes en filtros dependientes de la dirección y de la frecuencia llamados funciones de transferencia de cabeza (head-related transfer functions, HRTFs). Las HRTFs en audífonos dependen de la ubicación del micrófono (Algazi et al., 1999) y de la antropometría individual de cada persona (Algazi et al., 2001). En una síntesis binaural, si el conjunto de HRTFs utilizadas al filtrar los sonidos son las propias (HRTFs individuales), la fuente se percibirá de manera más compacta, externa y bien definida en una posición del espacio. En cambio, si las HRTFs pertenecen a otro individuo (HRTFs no individuales o genéricas) se producirá un deterioro en la capacidad de localización auditiva (Wenzel et al., 1993) y se incrementarán las confusiones delante-detrás (Bronkhorst, 1995; Middlebrooks, 1999).

Por otro lado, la inteligibilidad es la capacidad de reconocer las palabras contenidas en un fragmento de habla. En la literatura se han reportado beneficios relevantes de la audición binaural en la mejora de inteligibilidad en casos donde compiten ruido y habla de fuentes diferentes. En ese caso la audición espacial facilita la segregación y localización del habla (Avan et al., 2015).

La hipoacusia degrada las habilidades de localización espacial de los sonidos y la conservación de todas las claves de audición espacial es uno de los principales problemas que los audífonos actuales no logran resolver (Kollmeier & Kiessling, 2016). Las claves binaurales (ITD e ILD) pueden ser capturadas mediante el uso de dos audífonos sincronizados (conocidos como audífonos binaurales) uno para cada oído. Éstos permiten realizar la localización de sonidos a partir de la información complementaria de un procesamiento llamado beamforming (Widrow y Luo, 2003). Además, un beamformer binaural adaptativo puede mejorar la inteligibilidad de la palabra y reducir la carga cognitiva al escuchar (Picou et al., 2014).

Con respecto al ancho de banda de los audífonos, 8 kHz es suficiente para un correcto reconocimiento del habla. Sin embargo, se ha demostrado que el habla natural contiene información adicional entre 8 y 16 kHz que resulta esencial para lograr la correcta localización de los hablantes (Best et al., 2005). De esta forma, resulta necesario extender el rango normal de trabajo de los audífonos comerciales que varía entre los 4 y 6 kHz (Levy et al., 2015).

Por último, experimentos en escenarios de cocktail party -capacidad de focalizar la atención auditiva en un hablante particular dentro de un medio ruidoso- mostraron que el uso de información binaural en audífonos permite alcanzar mejoras en la calidad del habla y, por consiguiente de su inteligibilidad, hasta un 15% (Kavalekalam et al., 2019). También hay pruebas que señalan un beneficio adicional en la inteligibilidad si se tienen en cuenta características personalizadas del usuario de los audífonos (Moore et al., 2019)

El presente trabajo tiene como objetivo presentar una propuesta de investigación y desarrollo que se comenzó a realizar en el año 2021 y buscará realizar una contribución sobre aspectos aún no resueltos en los audífonos digitales actuales: mejora de la inteligibilidad del habla en ambientes sonoros complejos, y localización espacial de fuentes sonoras proporcionando una mejora en la naturalidad auditiva. Se plantea el desarrollo de un audífono binaural, el cual utiliza dos dispositivos -uno para cada oído- sincronizados, que permitirá incorporar las claves de audición binaural.

Desarrollo

Para lograr el objetivo planteado, se propone cumplimentar con cinco objetivos específicos:

Diseño de la etapa electroacústica de audífonos binaurales de múltiples micrófonos y altavoces sincronizables

El audífono binaural estará compuesto por dos dispositivos sincronizados que se integrarán a ambas orejas. Cada dispositivo, contendrá entre dos y cuatro micrófonos, y entre uno y dos altavoces balanceados. Se utilizará una combinación de micrófonos MEMS y electrets de tamaño reducido. Para el procesamiento de señales y la sincronización, se utilizará en una primera instancia, una computadora personal para realizar las pruebas off-line y, en una segunda instancia, un kit de desarrollo con microcontrolador de 32 bits. Las señales de los micrófonos serán digitalizadas mediante una placa de sonido diseñada ad-hoc, a una frecuencia de muestreo mayor a 32 kHz, logrando un ancho de banda de al menos 16 kHz y ampliando el rango de los audífonos convencionales. Esto permitirá incorporar claves monoaurales de alta frecuencia, críticas para la localización en el plano vertical y resolución de confusiones adelante/atrás.

Desarrollo de un prototipo inicial que integre procesamiento del habla para espacios sonoros complejos

Si bien el trabajo de investigación no se focaliza en desarrollar nuevas técnicas de análisis robusto, se propone implementar un conjunto básico de algoritmos y escenarios acústicos conocidos en la literatura, acordes a la capacidad de procesamiento del prototipo para su posterior evaluación en entornos reales. En un primer paso, se realizará el filtrado espectral, la cancelación de ruido impulsivo y el beamforming adaptativo (Van Veen, & Buckley, 1988). En un segundo paso, se implementarán técnicas de análisis de audio robustas al ruido: realce de habla (speech enhancement) y separación de fuentes sonoras (audio source separation), las cuales tienen un rol crítico en el desempeño de los dispositivos de mejora auditiva (Vincent et al., 2018). Como último paso, se propone emplear machine learning: redes neuronales artificiales para la detección de diferentes escenas auditivas.

Desarrollo de un modelo individualizado para mejorar el rendimiento en la localización sonora espacial

Idealmente, el diseño de la respuesta en frecuencia de los audífonos debería simular la función de transferencia individual a nivel de tímpano con el oído abierto del usuario (Denk et al., 2017), incorporando sus claves de localización. Se propone modelar una función de corrección individualizada de audífono para cada oído, que incluya las claves binaurales y monoaurales. A diferencia de Denk et al. (2018), será dependiente de la ubicación espacial de la fuente sonora y de la antropometría del pabellón auricular de cada sujeto. Esta función se deberá aplicar en tiempo real por el procesador del audífono binaural. Para su cálculo será necesario, en primer lugar, medir las HRTFs individuales del sujeto sin el audífono y con el audífono colocado. Se propone realizar estas mediciones en réplicas de las orejas a escala real. Estas réplicas impresas en 3D, ya se encuentran

construidas en el laboratorio a partir de una metodología para generar modelos antropométricos previamente elaborada por Tommasini et al. (2018). Serán colocadas en un HATS para obtener sus HRTFs mediante un sistema de medición desarrollado en el laboratorio (Guido et al., 2019; Scaliti et al., 2019). Estas mediciones realizadas sobre réplicas impresas no difieren significativamente de las mediciones en personas (Harder et al., 2016). Se medirán de esta manera a personas con audición normal, bajo dos condiciones: (a) sin el audífono, utilizando los micrófonos del HATS ubicados en la posición de los tímpanos (HRTF-i-sa), y (b) con el audífono colocado. Con estas HRTFs se calcularán las funciones de corrección individuales para todas las posiciones y se generará una base de datos que será embebida en el procesador. Contar con varios micrófonos por dispositivo permitirá aprovechar la información complementaria del beamforming para inferir la ubicación aproximada de la fuente sonora y seleccionar la función de corrección correspondiente de la base de datos para realizar un filtrado en tiempo real. Es necesario considerar que este modelado representa uno de los aspectos más importantes del trabajo, lo que permitirá realizar un aporte al área de estudio.

Evaluar objetivamente el prototipo inicial

En primer lugar, se evaluarán las diferencias existentes entre el conjunto HRTF-i-sa y dos nuevos conjuntos de mediciones utilizando el micrófono ubicado en el tímpano del HATS: (a) HRTFs utilizando réplicas individuales con audífono colocado y procesamiento de audición binaural activado (HRTF-i-ca-pa), y (b) HRTFs individuales con audífono y procesamiento desactivado (HRTF-i-ca-pd). Además, se realizará una medición de HRTFs genéricas, utilizando las orejas originales del HATS, con audífono colocado y procesamiento activado (HRTF-g-ca-pa).

En segundo lugar, se realizarán comparaciones con los conjuntos de mediciones de HRTFs a diferentes estilos de audífonos que se encuentran en base de datos públicas¹.

En ambos casos la evaluación se hará mediante métricas cuantitativas reportadas en la literatura. Mediante una interpretación conjunta de sus resultados, se intentará realizar una evaluación exhaustiva de las claves direccionales espaciales, y cómo son afectadas por el prototipo y el modelo que se propone.

Además, se evaluará de manera objetiva la inteligibilidad del habla en audífonos mediante métodos cuantitativos propuestos, por ejemplo, HASPI (Hearing Aid Speech Perception Index) y HASQI (Hearing Aid Speech Quality Index) (Kates et al., 2018).

Evaluar subjetivamente el prototipo inicial

Se propone realizar la validación subjetiva mediante dos pruebas psicofísicas, donde serán participantes aquellas personas a las cuales se les imprimió las réplicas de sus orejas.

Por un lado, se efectuará una prueba de inteligibilidad del habla que agregue realismo, para medir la tasa de reconocimiento de palabras, empleando una batería de evaluación de referencia de inteligibilidad de habla en ruido (Aronson et al., 2007). Además, se utilizarán diferentes grados de reverberación. Dichas condiciones son los escenarios más exigentes para reconocer el habla.

Por otro lado, para evaluar la naturalidad auditiva y la mejora en localización espacial, es necesario comparar directamente la percepción con el canal auditivo abierto y con el audífono binaural teniendo el procesamiento correspondiente aplicado. Se plantea la realización de una prueba de localización sonora espacial, en dos entornos diferentes: uno virtual y otro real. Para la prueba en entorno virtual se hará uso del sistema desarrollado por Tommasini et al. (2019). Se filtrarán los estímulos mediante diferentes conjuntos de HRTFs y se reproducirán mediante auriculares. Los participantes deberán enfrentar su nariz hacia dónde perciben el altavoz (virtual). La prueba en entorno real se realizará con múltiples altavoces ubicados en diferentes posiciones espaciales y sin auriculares. Para la condición control se usarán las HRTF-i-sa y para condiciones experimentales las HRTF-i-ca-pa, las HRTF-i-ca-pd y las HRTF-g-ca-pa. Se espera que en la condición con audífono y procesamiento activado se obtengan los resultados más similares a la condición control.

¹ <https://uol.de/mediphsyk/downloads/hearingdevicehrtfs>

Resultados y conclusiones

Se han realizado avances de este plan de investigación hasta la fecha. Fueron abordados cursos sobre análisis estadístico, diseño digital, metodología científica, microcontroladores y sistemas embebidos, procesamiento digital de señales, y redes neuronales. Se está realizando una revisión de la literatura científica sobre las características principales para la localización de fuentes sonoras en audífonos digitales binaurales. Actualmente, se encuentra en desarrollo la interfaz de audio de los audífonos en colaboración con otros miembros del equipo de investigación. Además se está investigando sobre la disposición espacial de los micrófonos sobre los cuales se basan los algoritmos de audición binaural.

Referencias

- Algazi, V. R., Avendano, C., & Thompson, D. (1999). Dependence of subject and measurement position in binaural signal acquisition. *Journal of the Audio Engineering Society*, 47(11), 937-947
- Algazi, V. R., Duda, R. O., Thompson, D. M., & Avendano, C. (2001, October). The cipic hrtf database. In *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics* (Cat. No. 01TH8575) (pp. 99-102). IEEE.
- Aronson, L., Estienne, P., Milone, D., Martínez, C., Rufiner, H., & Torres, M. E. (2007). Base de Datos Bateria de Evaluación para Pacientes con Prótesis Auditivas (BEPPA). *Editorial: FASO (Federación Argentina de Sociedades de Otorrinolaringología)*, 14(1), 17-24.
- Asano, F., Suzuki, Y., & Sone, T. (1990). Role of spectral cues in median plane localization. *The Journal of the Acoustical Society of America*, 88(1), 159-168.
- Avan, P., Giraudet, F., & Büki, B. (2015). Importance of binaural hearing. *Audiology and Neurotology*, 20(Suppl. 1), 3-6.
- Best, V., Carlile, S., Jin, C., & van Schaik, A. (2005). The role of high frequencies in speech localization. *The Journal of the Acoustical Society of America*, 118(1), 353-363.
- Bronkhorst, A. W. (1995). Localization of real and virtual sound sources. *The Journal of the Acoustical Society of America*, 98(5), 2542-2553.
- Ciorba, A., Bianchini, C., Pelucchi, S., & Pastore, A. (2012). The impact of hearing loss on the quality of life of elderly adults. *Clinical interventions in aging*, 7, 159.
- Denk, F., Ernst, S. M., Ewert, S. D., & Kollmeier, B. (2018). Adapting hearing devices to the individual ear acoustics: Database and target response correction functions for various device styles. *Trends in hearing*, 22, 2331216518779313.
- Denk, F., Hiipakka, M., Kollmeier, B., & Ernst, S. M. (2017). An individualised acoustically transparent earpiece for hearing devices. *International journal of audiology*, 57(sup3), S62-S70.
- Harder, S., Paulsen, R. R., Larsen, M., Laugesen, S., Mihocic, M., & Majdak, P. (2016). A framework for geometry acquisition, 3-D printing, simulation, and measurement of head-related transfer functions with a focus on hearing-assistive devices. *Computer-Aided Design*, 75, 39-46.
- Guido, R. M., Pucheta, M. A., Tommasini, F. C., Vergara, R. O., & Scaliti, F. (2019). Sistemas de Medición de HRTFS Individuales: Revisión del Estado del Arte y Desarrollos en Argentina. *Mecánica Computacional*, 37(5), 77-85.

- Kates, J. M., Arehart, K. H., Anderson, M. C., Muralimanohar, R. K., & Harvey Jr, L. O. (2018). Using objective metrics to measure hearing-aid performance. *Ear and hearing*, 39(6), 1165.
- Kavalekalam, M. S., Nielsen, J. K., Boldt, J. B., & Christensen, M. G. (2018). Model-based speech enhancement for intelligibility improvement in binaural hearing aids. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(1), 99-113.
- Kollmeier, B., & Kiessling, J. (2018). Functionality of hearing aids: State-of-the-art and future model-based solutions. *International journal of audiology*, 57(sup3), S3-S28.
- Levy, S. C., Freed, D. J., Nilsson, M., Moore, B. C., & Puria, S. (2015). Extended high-frequency bandwidth improves reception of speech in spatially separated masking speech. *Ear and hearing*, 36(5), e214.
- Moore, A. H., de Haan, J. M., Pedersen, M. S., Naylor, P. A., Brookes, M., & Jensen, J. (2019). Personalized signal-independent beamforming for binaural hearing aids. *The Journal of the Acoustical Society of America*, 145(5), 2971-2981.
- Picou, E. M., Aspell, E., & Ricketts, T. A. (2014). Potential benefits and limitations of three types of directional processing in hearing aids. *Ear and hearing*, 35(3), 339-352.
- Popelka, G. R., Moore, B. C. J., Fay, R. R., & Popper, A. N. (Eds.). (2016). *Hearing Aids. Springer Handbook of Auditory Research*. (S.I.): Springer International Publishing.
- Scaliti, F., Tommasini, F. C., Guido, R. M., Cravero, G. A., Capiglioni, L., & Guber, D. M. (2019). Amplificador de audio multicanal con control de salidas para aplicaciones experimentales de acústica. *Mecánica Comput.*, vol. 37, n.o 5, pp. 87-96, 2019.
- Tommasini, F. C., Cravero, G. A., Scaliti, F., Capiglioni, L., & Guido, R. M., (2018). Metodología para crear modelos antropométricos tridimensionales utilizados en la personalización de funciones de transferencia de cabeza. Congreso Latinoamericano de la AES, Montevideo, Uruguay. Montevideo, Uruguay, sep. 2018.
- Tommasini, F. C., Ramos, O. A., Hüg, M. X., & Ferreyra, S. P. (2019). A computational model to implement binaural synthesis in a hard real-time auditory virtual environment. *Acoustics Australia*, 47(1), 51-66.
- Van Veen, B. D., & Buckley, K. M. (1988). Beamforming: A versatile approach to spatial filtering. *IEEE assp magazine*, 5(2), 4-24.
- Vincent, E., Virtanen, T., & Gannot, S. (Eds.). (2018). *Audio source separation and speech enhancement*. John Wiley & Sons.
- Wallach, H. (1939). On Sound Localization. *The Journal of the Acoustical Society of America* (Vol. 10, Issue 4, pp. 270–274). Acoustical Society of America (ASA).
- Wenzel, E. M., Arruda, M., Kistler, D. J., & Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America*, 94(1), 111-123.
- Werner, L., Fay, R. R., & Popper, A. N. (2012). *Human Auditory Development*. Springer Handbook of Auditory Research. Springer New York.
- Widrow, B., & Luo, F. L. (2003). Microphone arrays for hearing aids: An overview. *Speech Communication*, 39(1-2), 139-146.