



Hybrid method for creating individual HRTFs CAD models

Sebastián Fingerhuth¹, Rodrigo Codoceo¹

¹Escuela de Ingeniería Eléctrica, Pontificia Universidad Católica de Valparaíso, Chile
sebastian.fingerhuth@pucv.cl

Resumen

En este trabajo presentamos los resultados de la creación de un modelo 3D CAD individualizado completo para aplicar a estudios de HRTF (Head Related Transfer Functions). Es un método híbrido que combina tanto un método para la *pinna* en base a moldes positivos y negativos y otro para la cabeza y torso en base a fotogrametría. El método de moldes genera un molde negativo de alginato de la *pinna* humana. Este molde es luego utilizado para crear un positivo de yeso para luego ser escaneado, incluyendo las concavidades, por ejemplo, la concha o el hélix.

Los modelos de cabeza y torso se obtienen en base a fotografías de la cabeza desde diferentes ángulos, generando una nube de puntos gracias.

Ambos modelos 3D CAD (mallas 3D) son combinadas usando puntos marcados.

Palabras-clave: HRTF, *pinna*, modelos 3D, fotogrametría, moldeo.

Abstract

In this paper we present the first results of creating a complete 3D CAD model of individual HRFT. It is a hybrid method that combines a molding method for the *pinna* and a photogrammetric 3D model of the head and torso. The molding method creates a alginate negative mold of the persons *pinna*. This mold then is used to create a positive plaster mold which can easily scanned, including concavities, for example in the cavum conchae or the helix.

The model of the head and torso is obtained from pictures taken from the head from different angles, combined to a point cloud in a photogrammetric software.

Both 3D CAD models (meshes) are combined using marked points.

Keywords: HRTF, *pinna*, 3D models, photogrammetry, molding.

PACS no. 43.64.+r

1 Introducción

La forma de la cabeza, el torso y la *pinna* (oreja) de los humanos juegan un papel importante en la localización espacial de una fuente sonora. Las reflexiones, especialmente en la *pinna*, que funcionan como filtro, están descritas por la función de transferencia relacionada a la cabeza (HRTF, Head Related Transfer Function). Estas funciones dependen de la posición de la fuente sonora y difieren entre cada oyente debido a la morfología propia de cada uno. Actualmente existen una gran serie de aplicaciones para HRTF desarrolladas en varios campos, por ejemplo, resulta de gran utilidad para aplicaciones médicas, grabación y reproducción 3D de sonido, teleconferencias para una realidad aumentada, cabinas



de pilotaje, videojuegos, etc. [1]. Las bases de datos para estas aplicaciones fueron obtenidas de varios sujetos pero no son completamente prácticas ya que para un uso más particular es necesario contar con HRTF individualizadas, es decir, obtenidas desde el individuo en particular que será usuario de la aplicación. Otra alternativa para encontrar HRTFs es mediante maniqués (*dummy head*) con medidas antropomórficas promedio o típicas. Estas entregan información de localización con un cierto error en la percepción de elevación, confusión en fuentes que están frente o atrás de la cabeza, entre otros. Una solución atractiva a estos antecedentes es el modelo individualizado de cabeza y *pinna* en 3D para posteriores mediciones acústicas.

El trabajo presentado acá se basa en una metodología de obtención de modelos 3D de cabeza y *pinna* individualizada para un sujeto. Esta metodología se divide en dos partes, la primera es la creación de las réplicas de la *pinna* mediante moldes negativos de alginato y posterior positivo de yeso para luego obtener a través de fotogrametría los modelos 3D CAD de la *pinna* derecha e izquierda. Y en segundo lugar es aplicar el trabajo fotogramétrico sobre la cabeza del sujeto para posteriormente unir todos los archivos 3D en un mismo cuerpo.

Es de importancia que los modelos 3D individualizados sean modelos exactos; se probaron y buscaron los parámetros y configuraciones ideales del equipamiento (cámara, óptica, software) que entregaron el mejor resultado en función a dimensiones morfológicas medidas en el individuo.

2 Antecedentes

Aquí se presentarán los conceptos teóricos referidos a la localización espacial de una fuente sonora, es decir, cuales son los fenómenos físicos que definen la ubicación de la fuente referidos al plano de coordenadas esféricas. También se dará introducción al software 3DF Zephyr para fotogrametría y los procesos internos que lo componen.

2.1 Conceptos básicos de espacialidad y direccionalidad del sonido

La HRTF describe las influencias que ejerce la anatomía de la cabeza y *pinna* en la percepción de la ubicación de la fuente sonora y la respuesta de amplitud y fase en función de la frecuencia; esto es, cómo cambia el sonido desde la fuente hasta llegar finalmente a ambos tímpanos [2]. Los cambios que sufre el sonido hasta llegar al oído pueden describirse como si la presión sonora estuviese bajo el efecto de un ecualizador de gran detalle que cambia sus características dependiendo de la dirección de la cual proviene el sonido [3][4].

Al localizar un sonido, la persona es capaz de ubicar el lugar de procedencia. Es práctico recurrir a las coordenadas esféricas para describir la posición. El sonido que llega a ambos tímpanos no es igual, sino que tiene diferencias que son principalmente las llamadas diferencias interaurales de intensidad y tiempo.

2.1.1 Diferencias interaurales

Las ILD (Diferencias interaurales de intensidad, *Interaural Level Differences*) se dan principalmente a partir de las diferentes distancias que deben recorrer las ondas para llegar a uno y otro oído, pero también por la sombra acústica producida por la cabeza del individuo (difracción de la onda). Una fuente sonora ubicada directamente al frente, atrás, arriba o abajo del oyente, será percibida con igual nivel de intensidad acústica por cada uno de los oídos y por lo tanto la diferencia es nula. Sin embargo si la fuente sonora es variada a la izquierda o derecha de alguna de estas cuatro posiciones iniciales, la percepción de nivel se inclinará hacia el oído que se encuentre más cercano a la fuente sonora [2].

Cuando el sonido llega a los oídos en momentos diferentes debido a que este recorre diferentes caminos, esta diferencia de tiempos es el ITD (Diferencias interaurales de tiempo, *Interaural Time Differences*).



Para frecuencias altas, la localización se debe fundamentalmente a la intensidad, mientras que para bajas frecuencias la localización que domina es la debida a la diferencia de fases y tiempos de retardos en el oído. Para frecuencias intermedias los seres humanos utilizan tanto la ILD como la ITD en forma combinada [2].

2.1.2 Ubicación en plano de elevación y distancia.

Mientras que la determinación del ángulo lateral es binaural (utiliza la información de ambos oídos), la determinación de la elevación es monoaural (un solo oído o sin diferencia entre los dos oídos). Es decir, se puede percibir la elevación con tan solo un oído. El efecto acústico del pabellón se basa en los fenómenos físicos de reflexión, difracción, dispersión, sombra acústica, interferencia y resonancia, y contribuye tanto a la correcta discriminación entre frente y atrás, como a la detección del grado de elevación de la fuente [2].

Para la ubicación espacial de la distancia de una fuente sonora se tienen en cuenta los siguientes mecanismos: intensidad, movimientos de la cabeza, exceso en las ILD y la razón entre el sonido directo y el reverberante.

2.2 Software 3DF Zephyr para fotogrametría.

La fotogrametría es una técnica para determinar las propiedades geométricas de los objetos y las situaciones espaciales a partir de imágenes fotográficas. Si se trabaja con una foto se puede obtener información bidimensional. Si se hace con dos fotos, en la zona común a éstas (zona de traslape), se puede tener visión estereoscópica, es decir, información tridimensional. La fotogrametría aplica para casi todas las cámaras, pero como las fotografías son resultado del equilibrio de varias variables, se utilizará una cámara DSLR (Digital Single Lens Reflex) ya que entrega la opción de contar con buena óptica (lentes) y la configuración manual de parámetros en la cámara.

El software 3DF Zephyr permite realizar modelos 3D fácilmente por medio de fotografías; el proceso es completamente automático una vez cargadas las imágenes y definidas las variables y parámetros del software. Para el desarrollo del mejor modelado 3D es necesario saber cómo trabaja y de qué forma se componen las tres tareas que conlleva el software para su objetivo, estas son: Structure from Motion, Multi-view Stereo y Mesh Extraction.

2.2.1 Structure from Motion.

Este campo de investigación se refiere a la recuperación de la geometría tridimensional de la escena (la estructura) cuando se observa a través de una cámara en movimiento (el movimiento). El resultado buscado es generalmente una nube de puntos 3D no estructurada, que consta de los puntos claves que fueron identificados en la escena y un conjunto de matrices de cámara, identificando la posición y dirección de cada imagen con respecto a un marco de referencia arbitrario.

2.2.2 Multi-view Stereo.

El objetivo de la Multi-view Stereo es extraer una densa reconstrucción de la superficie 3D a partir de múltiples imágenes tomadas desde puntos de vista de la cámara ya conocidos. El algoritmo se centra en la precisión, velocidad y escalabilidad [5]. Además, se puede integrar sin problemas como continuación del proceso anterior ya que se aprovecha de la nube de puntos escasa y su información de visibilidad.

2.2.3 Mesh Extraction.

Por último el proceso de obtención de la malla consta de unir por medio de segmentos los puntos generados en la nube densa producida en el algoritmo de multi-vista estéreo, el resultado son polígonos triangulares que forman una estructura sólida y final. El modelo 3D puede ser exportado a múltiples formatos para un uso posterior.

Finalmente los tres procesos, respetando tal secuencia, generan el camino a la construcción de un modelo escalado a proporciones reales del sujeto que sirvió de referencia, gracias a herramientas que el mismo software proporciona.

3 Modelo 3D para la cabeza.

Una parte importante del proceso es realizar una buena captura de imágenes, especialmente si se trata de cuerpos o partes humanas, por la gran cantidad de detalles definidos por la textura y las diferencias de profundidad. La idea principal es tomar las medidas para evitar que el modelo presente problemas. En el software se selecciona el modo *turntable*; la persona se sentará sobre una silla giratoria mientras la cámara permanece fija sobre un trípode variando solamente su altura.

Esta descripción es para la fotografía de la cabeza. Posteriormente también se fotografiarán los moldes de la orejas; por lo tanto es recomendable mantener las distancias cámara-objeto similar en ambos casos. Otro punto a considerar es que la densidad del cabello de la persona evita un reconocimiento fiel de los puntos clave a través de 3DF Zephyr, por tanto, un gorro de natación será necesario para evitar lo anterior y además mantener la forma del cráneo.

Es de importancia dejar marcadores en la cabeza y *pinna* de la persona para que luego el proceso de unión sea mucho más sencillo y además permite obtener un control de escalado de distancias real. En la Figura 1 se muestran algunos lugares claves para ubicar los puntos de control en la cara y alrededor de la *pinna*.

Mientras más marcadores, más pares de puntos serán unidos para hacer control de distancias y escalado, entonces la malla final de la cabeza será lo más cercana a las dimensiones reales.

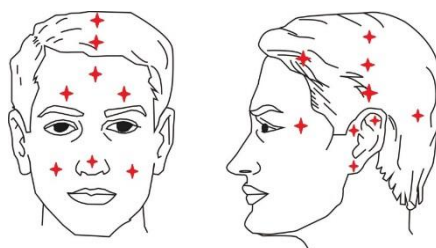


Figura 1 – Ubicación de marcadores o puntos de control

Hay que evitar poner marcadores en la zona baja de la cara, ya que, el movimiento ocasionado por la mandíbula durante la sesión fotográfica puede traer consecuencias que no favorecerán un correcto control de distancia y escalado de la malla final.

Gracias a la precisión de un pie de metro digital se miden las distancias de los marcadores en segmentos de distancia, y combinados entre sí. Mientras mayor es la cantidad de pares de marcadores registrados, mejor será el escalado, importante será entonces cubrir la mayor superficie de la cabeza con estas distancias medidas.

La sesión fotográfica consta de una órbita circular alrededor de la cabeza y otra serie de fotografías que cubran el sector bajo el mentón y con más dedicación sobre el lóbulo parietal para que resulte una malla 3D cerrada; esta sesión debe ser de técnica *turntable*.

El escenario fotográfico dispuesto para esta parte del procedimiento se muestra gráficamente en la Figura 2, y debe satisfacer varios puntos:

- Una apertura de diafragma que enfoque bien los bordes de la cabeza (nariz, barbilla, *pinnas*, etc.).
- Normalmente la poca luz que entra al sensor por la baja apertura (A_v) es equilibrada con un tiempo de exposición mayor (T_v). Este último factor puede ocasionar una imagen borrosa por el movimiento de la persona mientras se toma la fotografía. Es necesario entonces tener fuentes de luz externa (flash, reflectores, etc.) y reducir el tiempo.
- Como se ha hecho hincapié anteriormente, mantener siempre el lente de la cámara abarcando todo el objetivo; usar una distancia focal suficiente para evitar la distorsión de la luz que ingresa y no deformar la cabeza.
- La cámara cuenta con nueve puntos de enfoque. Es recomendable que el objetivo deba permanecer dentro de estos nueve puntos para que salga nítido en todos sus bordes, así se aprovecha completamente el ajuste hecho a la apertura combinado con una distancia focal de 35 [mm].

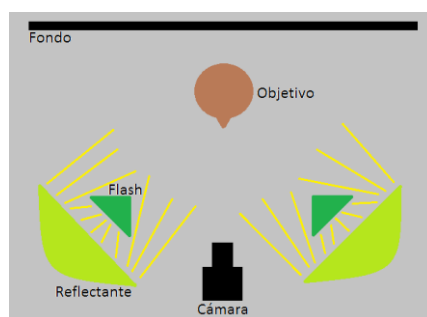


Figura 2 – Disposición de la iluminación en el escenario fotográfico.

Los parámetros para la sesión fotográfica que se muestra en la Tabla 1 se utilizan para comenzar con el modelado 3D de la cabeza.

Tabla 1 – Parámetros de cámara para sesión fotográfica de cabeza.

N° de Fotografías	Tiempo de exposición [s]	Apertura Diafragma	ISO	Distancia focal [mm]	Distancia al objetivo [cm]	Flash
24	1/200	f/16	400	35	200	Sí

Una vez obtenidas las imágenes, se procesan en 3DF Zephyr. Conseguido el modelo 3D los marcadores colocados en las orejas deberán ser visibles con el fin de adherir más tarde las *pinnas* izquierda y derecha a la cabeza mediante el proceso de *Workspace Merging* de Zephyr.

Los puntos de control obtenidos se colocan sobre el modelo 3D de cabeza y se seleccionan todos los pares de puntos de los cuales será necesario generar el segmento de distancia previamente medido. La Figura 3 muestra la etapa descrita anteriormente. Se mostrarán de manera automática las distancias actuales que posee la malla luego de agregar cada segmento por par de puntos; por consiguiente, se deben incluir las distancias reales para poder escalar la malla.

El escalado realizado fue gracias a la medida de 60 segmentos, con todas estas distancias agregadas el software acomoda la malla generando un error de longitud promedio de 3.845 [mm] equivalente a un error promedio de 2.33%.

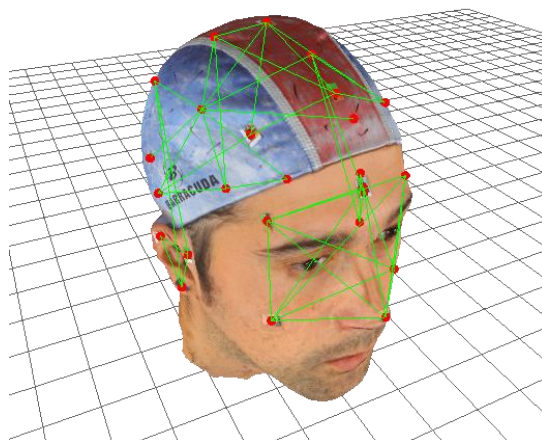


Figura 3 – Segmentos de distancia en malla 3D de cabeza.

4 Modelo 3D para la *pinna*.

Posterior a la toma de fotografías de la cabeza, se crean los moldes de alginato para las *pinnas*, izquierda y derecha. Cabe recordar que en cada una de ellas se colocaron marcadores. Es necesario evitar que se muevan o se pierdan *durante* el proceso. Se creó un molde que contenga la *pinna* del sujeto en el cual se vertió el alginato en forma líquida. Una vez que el alginato endureció se retira de la *pinna* original y se procede a seccionar el negativo por un corte de cuchillo. Así es posible presentar/ exponer los espacios más ocultos (concavidades). En la Figura 4 se muestra el resultado.

Luego se fabrica el molde positivo en yeso. Una vez secado el yeso, se retira el alginato, se eliminan las imperfecciones en el yeso ocasionadas por burbujas de alginato (si las hay) [3], y se marcan las réplicas para generar los puntos de interés.

En este proceso se pueden encontrar algunos problemas, uno de ellos es que la *pinna* que sirve de modelo puede tener algunos canales muy estrechos (dependiendo de la persona), específicamente donde convergen la *cymba* y la fosa triangular, lo que ocasiona que si el alginato no es lo suficientemente líquido, no alcanza a cubrir esta zona. Para solucionar esto, al agregar la mezcla líquida a la contención se puede ayudar con los dedos para que el fluido llegue a las zonas más difíciles, obteniendo un negativo más detallado.

El yeso utilizado es de estilo ortopédico, su característica es que es mucho más blando que el convencional. Así es más sencillo de modelar con herramientas la base de la *pinna* de yeso que se creó. El fin es exponer la zona posterior al pabellón auricular y así hacer más sencillo obtener información de ese lugar en el proceso de la fotografía de las *pinnas*. En algunas personas el pabellón auricular está muy junto al cráneo, debido a esto se toma la medida anterior.



Figura 4 – Proceso de creación de los negativos de alginato de *pinnas* izquierda y derecha.

4.1 Mallas 3D de *pinnas* izquierda y derecha.

Para las *pinnas* izquierda y derecha se realiza el procedimiento para cada par de trozos de ellas del mismo modo. Con cada trozo sobre el pedestal, tres órbitas fotográficas y con parámetros de cámara de la Tabla 2 se toman las fotografías.

Finalmente se obtienen los 4 trozos de *pinna* en mallas 3D. El proceso para crear la *pinna* completa se mostrará a continuación, en este caso será la oreja derecha. Para la izquierda se sigue el mismo procedimiento.

En la cara interna del trozo A y B (Figura 4) de la oreja derecha se seleccionan los puntos de control en común para ambos, estos puntos se eligen aleatoriamente o si se hicieron marcas para una mejor guía de unión es más favorable. Los puntos seleccionados en cada trozo deben mantener además de la ubicación en común, el nombre de ellos. Estos puntos están nombrados: 'W', 'X', 'Y' y 'Z'.

Tabla 2 – Parámetros de cámara para sesión fotográfica de las *pinnas*.

Nº de Fotografías	Tiempo de exposición [s]	Apertura Diafragma	ISO	Distancia focal [mm]	Distancia al objetivo [cm]	Flash
33	1/8	f/16	800	50	15	No

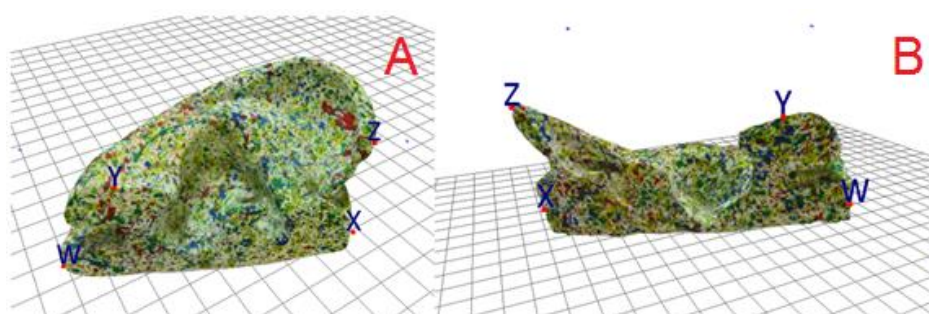


Figura 5 – Malla de trozos A y B, y marcadores para la unión.

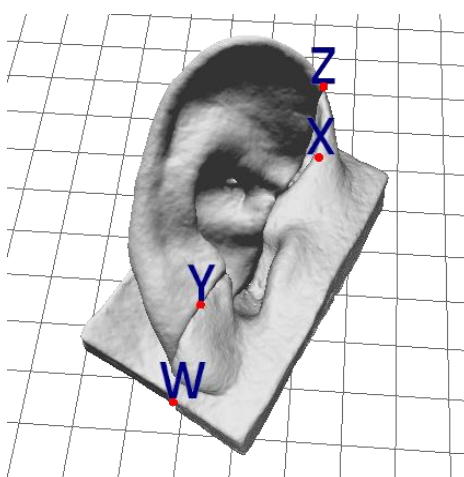


Figura 6 – *Pinna* derecha reconstruida.

Siguiendo con el proceso se crea el modelo 3D de la *pinna* izquierda y nuevamente se debe escalar el modelo. El marcar con puntos de colores las réplicas de yeso genera una gran cantidad de puntos de control de distancias. Por lo tanto se puede escalar con gran precisión la *pinna*. Esto es de gran importancia para las futuras mediciones acústicas de HRTF.

El error medio porcentual luego del escalado fue de 1.67% (0.4134 [mm]) para la *pinna* derecha y de 1.11% (0.3372 [mm]) para la izquierda, valores muy pequeños que son aceptables e indican que los modelos 3D de *pinna* se acercan lo suficiente a las dimensiones reales.

5 Unión final de *pinnas* y cabeza.

La malla 3D que resultó de la cabeza tenía las orejas del individuo con mínimo detalle pero mantenía los marcadores intactos, por tanto el proceso a seguir es seleccionar en aquella malla los puntos de control y seleccionar los marcadores en el sector de la *pinna* (unir una a la vez). Al crear las réplicas de yeso de ambas *pinnas*, los marcadores generaron un pequeño relieve sobre la superficie, por tanto es sencillo ubicar los puntos de control sobre la *pinna*. Se importa uno de los archivos al complementario espacio de trabajo y se unen por medio de un proceso interno en 3DF Zephyr. Se repite el proceso para la otra *pinna*. Las *pinnas* detalladas se verán combinadas con las que resultaron del modelo 3D de la cabeza, las cuales hay que eliminar.

El espacio de trabajo estará compuesto por cinco mallas: dos trozos por cada *pinna* y la malla de la cabeza como se muestra en la Figura 7, esto genera gran consumo computacional por la cantidad de cámaras y nube de puntos que se han adicionado al avanzar el proceso. Finalmente las mallas son fusionadas gracias a otro software adicional y así crear una malla poligonal única detallada (Figura 8).

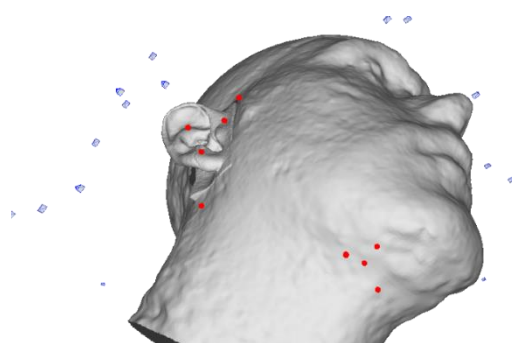


Figura 7 – Unión de *pinnas* y cabeza mediante marcadores en 3DF Zephyr.



Figura 8 – Malla final de cabeza y *pinnas* individualizadas.

Finalmente se ha obtenido una malla 3D de cabeza, *pinna* izquierda y derecha en alta definición., este método es una de tantas formas de escaneo individualizado. Existen más alternativas que varían en tecnología y metodología [6].

6 Conclusiones.

En el trabajo se hicieron pruebas con distintas configuraciones, tanto de la óptica y cámara, como de los parámetros configurables del software.

Se verificó que con tres órbitas (horizontal, superior e inferior) alrededor de las *pinnas* entregaba la cantidad necesaria de capturas para procesar en 3DF Zephyr. Para la sesión fotográfica de la cabeza basta con una órbita media a la altura de los ojos de la persona y fotografías adicionales sobre la cabeza y bajo el mentón para cubrir toda la superficie.

En este trabajo se ha presentado una metodología híbrida para obtener modelos 3D de cabeza y *pinnas* individualizadas con el fin de obtener HRTFs individualizadas. Los análisis de los resultados, que es comparar distancias reales con las distancias en los modelos CAD indican que el proceso de alta precisión, con errores cercanos al 2%.

Para resolver el problema para obtener los modelos 3D de las *pinnas*, que incluye zonas ocultas a la fotografía (concavidades) se cortó el molde negativo de alginato, antes de verter el yeso. Posteriormente se recrearon de forma separada los trozos que componían a la *pinna*



Dentro del uso del software los procesos más importantes fueron el escalado de los modelos a las dimensiones reales y la unión de ellos para formar la cabeza completa con *pinnas*. El proceso de escalado resulta más eficiente si se respetan desde un principio los parámetros de cámara, sesión fotográfica y parámetros de software para la reconstrucción; estos modelos entregan distancias reales con mínimo porcentaje de error si se utiliza gran cantidad de segmentos de distancia como restricciones para el escalado final. Las distancias de error mínimas para los modelos 3D de la *pinna* son suficientemente pequeñas para no significar problema a las longitudes más cortas de onda sonora audible, es decir, que son aptas para mediciones acústicas entre 20 [Hz] y 20 [kHz].

El proceso de unión de los modelos 3D fue realizado con la ayuda de marcadores ubicados sobre la *pinna* de la persona y luego considerados en el software como puntos de referencia para la unión. La unión final entrega como resultado una malla 3D de cabeza y oreja individualizada exportable en múltiples formatos para cualquier uso posterior, ya sea, impresión 3D o simulaciones acústicas para la obtención de la HRTF.

7 Bibliografía

- [1] Craig T. Jin, Pierre Guillon, Nicolas Epain, Reza Zolfaghari, and André van Shaik, "CREATING THE SYDNEY YORK MORPHOLOGICAL AND ACOUSTIC RECORDINGS OF EARS DATABASE," *IEEE TRANSACTIONS ON MULTIMEDIA*, vol. 16, no. 1, pp. 37-46, Enero 2014.
- [2] Nicolás Maserá, Juan Morales, and Ana Moreno, "Claves perceptuales sobre direccionalidad y espacialidad del sonido" Documento de Cátedra Fundamentos de Acústica y Electroacústica UTN, Córdoba, Argentina, 2011.
- [3] José Luis Bravo, "Construcción de modelo de oreja artificial de silicona y medición de características acústicas," Tesis de proyecto de título en Ingeniería Electrónica, PUCV, Valparaíso 2014.
- [4] Janina Fels, "From children to adults: how binaural cues and ear canal impedances grow," RWTH Aachen University, Berlin, Tesis doctoral 2008.
- [5] Roberto Toldo, "Towards automatic acquisition of high-level 3D models from images," PHD Thesis, Verona, Italia, 2013.
- [6] Andreas Reichinger, Piotr Majdak, Robert Sablatnig, and Stefan Maierhofer, "Evaluation of Methods for Optical 3-D Scanning of Human *Pinnae*," in *2013 International Conference on 3D Vision*, Viena, Austria, 2013, pp. 390-397.