ESPACIO ACÚSTICO VIRTUAL -EAV- A PARTIR DEL PROCESAMIENTO Y ANÁLISIS DIGITAL DE IMÁGENES DE MAPAS DE PROFUNDIDAD

IVÁN DARÍO CORCHUELO CASTRO

UNIVERSIDAD DEL CAUCA

FACULTAD DE CIENCIAS NATURALES, EXACTAS Y DE LA EDUCACIÓN

DEPARTAMENTO DE FÍSICA

INGENIERÍA FÍSICA

POPAYÁN

2013

ESPACIO ACÚSTICO VIRTUAL -EAV- A PARTIR DEL PROCESAMIENTO Y ANÁLISIS DIGITAL DE IMÁGENES DE MAPAS DE PROFUNDIDAD

INFORME DE TRABAJO DE GRADO PRESENTADO COMO REQUISITO PARA OPTAR AL TITULO DE INGENIERO FÍSICO
Modalidad de Investigación

IVÁN DARÍO CORCHUELO CASTRO

Director
Leonairo Pencue-Fierro

UNIVERSIDAD DEL CAUCA
FACULTAD DE CIENCIAS NATURALES, EXACTAS Y DE LA EDUCACIÓN
DEPARTAMENTO DE FÍSICA
INGENIERÍA FÍSICA
POPAYÁN
2013

Nota de aceptación

_____

_____

_____

_____

_____

Ing. Leonairo Pencue Fierro
Director

_____

Dr. Rubiel Vargas Cañas
Evaluador del Proyecto

_____

Mg. Carlos Felipe Ordoñez
Evaluador del proyecto

Fecha de sustentación: Popayán, Abril 26 de 2013

AGRADECIMIENTOS

Agradezco a Dios por su gran bondad, misericordia y gracia salvadora, por El tengo vida para brindar servicio en mi profesión.

A mis padres, Miguel y Claudia, por brindarme aliento y apoyo, seguridad y firmeza durante este camino universitario. Ellos me han dado de sus virtudes y principios para vivir. A mis hermanos, Diego, David, Daniel y César, que han sido ejemplos de vida, trabajo y estudio. De ellos, he reconocido el esfuerzo, el valor y la hermandad.

A mi novia y amigos, Viviana, Tatiana, David, Laura, Francia, William, Wilmer, por ser personas especiales en los caminos universitario y colegial, disfrutando de buenos y agradables momentos. A mis hermanos de Unidad Cristiana Universitaria (UCU), Guillermo, Káterin, leidy, Johny, Ángela, Grace, Bertica, por darme palabras de sabiduría y esperanza. A la Iglesia Cristiana Cuadrangular de Popayán y los pastores, por su acompañamiento pastoral.

A la Universidad del Cauca, profesores del Departamento de Física, a mi director Leonairo Pencue, a Mario Patiño, a los jurados Rubiel Vargas y Carlos Ordoñez, por sus colaboraciones y enseñanzas. A mis compañeros de la universidad, Carlos (peke), Laura, Luza, Julio, Adriana, Pablo, Guillermo, Jazmín, Karen, Andrea, Santiago, por tantos momentos agradables y llenos de alegría.

Agradecimientos especiales al profesor Hernando Silva y a la profesora Ana Sol Restrepo por su colaboración en las investigaciones del proyecto, y en general a todas las personas que estuvieron a mi lado durante este proceso lleno de satisfacción y de buenos recuerdos de las cuales pude aprender en mi caminar universitario.

CONTENIDO

pág.

LISTA DE FIGURAS

pág.

LISTA DE TABLAS

pág.

INTRODUCCIÓN

En situaciones de limitación visual o ausencia de luz visible, una persona presenta dificultades para percibir la posición de objetos en su entorno, la situación es más compleja cuando aquellos no producen sonido alguno y están fuera del alcance del tacto. A partir de esto el ser humano puede desarrollar habilidades para mejorar la percepción de otros sentidos, como la audición, y apoyarse en la flexibilidad del cerebro humano para realizar una sustitución sensorial de las funciones visuales deterioradas, y así poder desempeñarse en actividades cotidianas.

Esta investigación presenta una alternativa para asistir a las personas en la ubicación de objetos en su entorno por medio del sentido de la audición. Se desarrolla un sistema que reproduce un Espacio Acústico Virtual (EAV) a partir de imágenes de mapas de profundidad utilizando Funciones de Transferencias Acústicas y herramientas software disponibles especialmente en los campos de visión por computador y sonidos 3D. Para obtener las imágenes de los mapas de profundidad se utiliza el sensor de profundidad Kinect.

La investigación presentada en este documento se ha estructurado de la siguiente manera: en el capítulo 1 menciona el principal sector de aplicación al que está dirigida la investigación, sin descartar otras posibles aplicaciones y sectores donde pueda ser usado el sistema. El capítulo 2 realiza una breve descripción de otros sistemas que proponen alternativas de solución similares que se tuvieron en cuenta para el desarrollo del proyecto. Los capítulos 3 y 4 exponen los fundamentos teóricos sobre el Espacio Acústico Virtual y el sistema de visión por computador a implementar, además de las consideraciones que debe tener el sistema según el sistema de audición humano.

En el capítulo 5, se presenta el diseño y la implementación del sistema en tres etapas descritas mediante esquemas y diagramas de flujo: 1) adquisición de los mapas de profundidad, 2) procesamiento y análisis de los mapas de profundidad, y

3) virtualización y reproducción del Espacio Acústico Virtual. Adicionalmente se mencionan las Funciones de Transferencia Acústicas involucradas en el proceso de sintetización del sonido, así como del dispositivo óptico utilizado.

En el capítulo 6, se muestran los resultados de las etapas y de las pruebas de desempeño del sistema con su respectivo análisis. Para los resultados de las etapas, se realiza un ejemplo del funcionamiento del sistema con una imagen de mapa de profundidad que permite obtener la información que cada etapa entrega. En cuanto a las pruebas, se presentan las tablas de resultados de la detección de objetos respecto a su ubicación, cantidad y orden en el que se encuentran.

Finalmente, se mencionan las conclusiones que esta investigación arroja con base en los resultados obtenidos y se exponen los posibles trabajos futuros que darían continuidad a la investigación en busca de alternativas de solución complementarias.

1. DISCAPACIDAD VISUAL.

Durante la percepción, el sentido de la vista le aporta a un individuo abundante información sobre la ubicación de los objetos en un determinado lugar, frente a la que se pueda percibir a través de los otros sentidos. Es así como puede determinar el número de objetos presentes en un escenario y la distancia a la que se encuentran.

Cuando la función visual presenta deficiencias y limitaciones, los procesos de interacción con el entorno se dificultan, por lo tanto el individuo acude a otros mecanismos sensoriales para identificar la posición de los objetos.

1.1 Dimensiones de la situación actual.

Conforme a la Clasificación Internacional de Enfermedades (CIE-10, actualización y revisión de 2006), la función visual se subdivide en cuatro niveles:

- Visión normal
- Discapacidad visual moderada
- Discapacidad visual grave
- Ceguera.

Existen diversas causas que degeneran la función visual en un individuo y dificultan la ubicación de objetos y obstáculos en el entorno, entre ellas están los errores de refracción (miopía, hipermetropía o astigmatismo), cataratas, glaucoma, diabetes no controlada, oncocercosis y tracoma. Según la Organización Mundial de la Salud (OMS), en el mundo hay aproximadamente 285 millones de personas con Discapacidad Visual (DV), de las cuales 39 millones son invidentes y 246 millones presentan baja visión. En la Tabla 1 se presentan las principales causas que producen la discapacidad visual reportadas por la OMS en el mundo (ORGANIZACIÓN MUNDIAL DE LA SALUD, 2012).

Tabla 1. Principales causas de la discapacidad visual.[1]

Causas	Población con DV (%)
Errores refractivos no corregidos (miopía, hipermetropía y astigmatismo)	43
Cataratas	33
Glaucoma	2
Otros (...)	22

El 80% de los casos en las personas con ceguera son evitables o susceptibles de tratamiento. Si la causa es tratada debidamente, puede progresar y llegar a condiciones óptimas en la función visual. Por eso, la OMS adelanta campañas en la implementación de programas para la prevención y el control de la discapacidad visual. Actualmente se está desarrollando el *Action plan for the prevention of avoidable blindness and visual impairment for 2014-2019*[2]

En Colombia se estima que el 2.8% de la población presenta algún tipo de Discapacidad Visual. Según estadísticas del Plan Nacional de Atención a las personas con discapacidad PNAD (2006), DANE, MEN y las Administraciones Municipales y Departamentales, existe un total de 283.726 personas con una o más limitaciones; de ellas, 87.525 presentan una deficiencia visual; es decir, un problema en la función o estructura visual, como una desviación o una limitación de acuerdo con la Clasificación Internacional del Funcionamiento, de la Discapacidad y de la Salud (CIF). De este grupo 38.130 (13.4%), son personas con graves problemas visuales que les ocasiona discapacidad visual; por tanto, presentan limitaciones en las actividades cotidianas y restricciones en la participación social (INSTITUTO NACIONAL PARA CIEGOS, 2006).

La Tabla 2 agrupa la situación de discapacidad en Colombia por departamentos. Indica el número de personas registradas con alguna de las limitaciones visuales y

[1] OMS, http://www.who.int/mediacentre/factsheets/fs282/es/index.html
[2] OMS, http://www.who.int/blindness/actionplan/en/index.html

el número de personas con limitación visual completa o ceguera, resultado del censo del DANE en el año 2005.

Tabla 2. Limitación y discapacidad visual en Colombia[3]

Departamento	Registrados con alguna limitación visual	Registrados con limitación visual	%
Antioquia	21.420	2.491	11,6%
Atlántico	14.413	1.883	13,1%
Bogotá	49.947	4.328	8,7%
Bolívar	515	51	9,9%
Boyacá	11.653	1.528	13,1%
Caldas	1.323	216	16,3%
Casanare	6.308	1.008	16,0%
Cauca	**21.239**	**2.664**	**12,5%**
Cesar	15.296	2.931	19,2%
Córdoba	24.794	4.281	17,3%
Cundinamarca	7.720	1.022	13,2%
Huila	25.994	3.907	15,0%
La Guajira	8.987	1.784	19,9%
Nariño	14.352	1.529	10,7%
Santander	3.897	498	12,8%
Tolima	31.211	5.143	16,5%
Valle	24.657	2.866	11,6%
Total	283.726	38.130	13,4%

En la Tabla 3 se registra la distribución de la población con limitación visual en el departamento del Cauca, desde el censo del 2005 con corte en marzo del 2009. Con base en dicha información, se observa un incremento del 26,35% en el registro de la población con limitación visual en el departamento del Cauca en el periodo 2005-2009. Se espera que esta cifra no siga en aumento, razón por la cual las entidades que les corresponde atender esta situación, continúan adelantando actividades de prevención y rehabilitación.

[3] INSTITUTO NACIONAL PARA CIEGOS. (Octubre de 2006). *Estadisticas de Discapacidad Visual en Colombia.*

Tabla 3. Limitación y discapacidad visual en el Cauca[4]

MUNICIPIO	TOTAL	%	MUNICIPIO	TOTAL	%
ALMAGUER	84	2,50	PÁEZ	109	3,24
ARGELIA	54	1,60	PATÍA	51	1,52
BALBOA	19	0,56	PIAMONTE	31	0,92
BOLÍVAR	218	6,48	PIENDAMÓ	93	2,76
BUENOS AIRES	89	2,64	POPAYÁN	551	16,37
CAJIBÍO	5	0,15	PUERTO TEJADA	120	3,57
CALDONO	23	0,68	PURACÉ	11	0,33
CALOTO	65	1,93	ROSAS	40	1,19
CORINTO	28	0,83	SAN SEBASTIÁN	67	1,99
EL TAMBO	176	5,23	SANTA ROSA	112	3,33
FLORENCIA	39	1,16	SANTANDER DE Q/CHAO	204	6,06
GUAPÍ	246	7,31	SILVIA	19	0,56
INZÁ	77	2,29	SOTARÁ	55	1,63
JAMBALÓ	15	0,45	SUÁREZ	82	2,44
LA SIERRA	26	0,77	SUCRE	7	0,21
LA VEGA	92	2,73	TIMBÍO	85	2,53
LÓPEZ DE MICAY	39	1,16	TIMBIQUÍ	139	4,13
MERCADERES	66	1,96	TORIBÍO	39	1,16
MIRANDA	32	0,95	TOTORÓ	29	0,86
MORALES	29	0,86	VILLA RICA	36	1,07
PADILLA	64	1,90	**Total**	**3.366**	**100,00**

En el departamento del Cauca, se encuentran dos redes de trabajo frente a la situación de discapacidad visual, la Red Juntos y la Red Pensar. La Red Juntos es una estrategia de intervención integral y coordinada de los diferentes organismos y entidades del Estado, que tiene por objeto mejorar las condiciones de vida de las

[4] Correa, L. M. (2012). *Documento Territorial Cauca N° 10.*

familias en situación de pobreza extrema y lograr que estas familias puedan generar sus propios ingresos de manera sostenible. Y la Red Pensar trabaja directamente con la población en situación de discapacidad, tiene convenios con las entidades: instituto Colombiano de Bienestar Familiar (ICBF), Caja de Compensación Familiar del Cauca (COMFACAUCA), Hospital Nivel II Susana López de Valencia y la Gobernación del Cauca (Correa, 2012).

En la ciudad de Popayán se presentan acciones de atención a la población con limitación visual en el área educativa, laboral, de formación y capacitación, rehabilitación funcional, materiales y equipos de donación y apoyo. Sin embargo, en el campo de la investigación en el tema de la salud visual, a la fecha no se reportan trabajos.

En Colombia las principales entidades que apoyan a la población en situación de discapacidad visual son: el Instituto Nacional para Ciegos (INCI), que brinda asesoría y asistencia técnica a entidades de nivel nacional y departamental, esta entidad propone políticas, planes y programas para mejorar la calidad de vida de personas en situación de discapacidad visual, ceguera y baja visión; y la otra entidad importante es el Centro de Rehabilitación para Adultos Ciegos (CRAC), una fundación de carácter privado encargada de facilitar la inclusión social de personas en situación de discapacidad visual. Además, establece políticas y desarrolla programas indispensables para mejorar la calidad de vida de las personas mediante el aprendizaje y/o adquisición de habilidades y destrezas que les permitan aportar al desarrollo económico y social de su comunidad.

1.2 Entrenamiento en Orientación y Movilidad

Las personas que tienen una discapacidad visual pueden recibir tratamientos, desde el uso de lentes correctivos hasta operaciones quirúrgicas. No siempre todos los tratamientos corrigen completamente la discapacidad visual. En situaciones donde son difícilmente aplicables por razones médicas, económicas, culturales,

entre otras, y la discapacidad visual es muy grave (ceguera), es factible iniciar un proceso de rehabilitación que incluye un entrenamiento en Orientación y Movilidad (O&M) (Sáez, 2010).

El entrenamiento en O&M posibilita al individuo aprender a organizar, familiarizarse e interactuar con su entorno. La orientación trata de hacer comprender al individuo quién es, dónde está y hacia dónde se quiere desplazar; mientras que la movilidad corresponde al acto del desplazamiento.

Existen varios conceptos en la O&M que es necesario explorar durante el entrenamiento:

- **Imagen Corporal**: Es el concepto sobre su propio cuerpo.
- **Concepto Corporal**: Conocer las partes del cuerpo y su funcionalidad
- **Conciencia Sensorial**: Comprender que se recibe información del entorno por medio de los sentidos.
- **Conceptos Espaciales**:
    - **Permanencia de Objetos**: Comprender que los objetos existen aún cuando no se puedan percibir (escuchar, tocar, oler).
    - **Nociones Espaciales**: Entender las relaciones espaciales existentes entre los objetos, como arriba, abajo, cerca, lejos, delante, atrás, entre, dentro, fuera.
    - **Nociones Temporales**: Comprender que existe ayer, hoy y mañana, que hay día y noche, semanas, meses y años.
- **Capacidad de Búsqueda**: Aprender a buscar objetos y encontrarlos.
- **Movimientos Independientes**: Capacidad de girar, rodar, gatear, caminar.
- **Guía Vidente**: Usar el apoyo de una persona que ve para desplazarse.
- **Técnicas de Protección**: Aprender destrezas que se usan en situaciones específicas de desplazamiento y sirven para protección.

Entre los servicios que presta el CRAC, se encuentran los apoyos pedagógicos especializados en áreas tiflológicas, como la orientación a niños, niñas y jóvenes con discapacidad visual para la integración en el aula regular con edades entre cinco y catorce años, y la enseñanza en áreas tiflológicas (braille, ábaco, elementos de comunicación, orientación, movilidad y sensopercepción, Actividades Básicas Cotidianas ABC).

Al finalizar el entrenamiento, el individuo aumenta el índice de calidad de vida a través de un mayor grado de independencia en tareas cotidianas, favoreciendo su interacción en ambientes familiares y la integración con la sociedad.

Las personas en situación de discapacidad visual constituyen el sentido del presente proyecto de investigación, fundamentalmente en términos de orientación y movilidad, a través del estudio de una alternativa para recibir información sobre la ubicación espacial de objetos en el entorno. Para esto, los conceptos espaciales y la conciencia sensorial del entrenamiento en O&M son importantes para la construcción del Espacio Acústico Virtual. La permanencia de los objetos, la noción espacial y la ubicación de objetos son aspectos que se desarrollan en detalle más adelante.

1.3 Alternativas de aplicación

El EAV presenta múltiples formas de uso y aplicaciones. Además de la O&M, es posible brindar asistencia en procesos educativos, apoyando la inclusión de los estudiantes con limitaciones visuales en las instituciones educativas. Puede asistir en las formas de aprendizaje y desarrollo de nuevos conceptos a partir de sonidos virtuales y la interacción de los mismos para recrear un escenario mental.

El sector de entretenimiento es un gran campo de aplicación para el proyecto de investigación, la generación de un EAV en contacto con espacios reales, permitiría incursionar en el aspecto de la realidad virtual. Reproducir sonidos de objetos

virtuales en ambientes reales aumenta los efectos de la realidad virtual aumentada (Wikipedia, Realidad aumentada, 2013) y permite desarrollar interacciones más naturales (OpenNI, 2010)

En el campo de la medicina, permitiría realizar evaluaciones del sistema auditivo, en cuanto a la estereofonía, agudeza auditiva, sensibilidad audible y ubicación de objetos.

2. TIFLOTECNOLOGÍA

La tiflotecnología es una tecnología para la asistencia a personas en situación de discapacidad visual que permite incrementar, mantener y mejorar las capacidades funcionales de las personas. La principal función tratada en el desarrollo del proyecto es la capacidad de ubicar objetos en el entorno.

Diversos dispositivos permiten la posibilidad de ubicar objetos en el espacio de diferentes formas y estímulos sensoriales, como señales acústicas para el sistema auditivo, impulsos eléctricos para la sensibilidad en la lengua y vibraciones mecánicas para sensaciones en la piel. Entre los dispositivos a mencionar están BrainPort (Bach-y Rita, *et al*, 1998), CASBLiP (Praderas, *et al*, 2009), VoiCe (Meijer, 1992) y NAVI (Zöllner, *et al*, 2011).

2.1 BrainPort

Este dispositivo se basa en las investigaciones de Paul Bach-y Rita (Bach-y Rita, *et al*, 1969) sobre la sustitución sensorial, en el que la plasticidad neuronal permite realizar funciones visuales sin recibir la información espacial propiamente del sentido de la vista, es decir que el ser humano *ve* con el cerebro, no con los ojos.

En el caso del dispositivo BrainPort la función visual se realiza por medio de las terminales nerviosas de la lengua. Consta de una cámara situada en la cabeza de la persona sobre un par de gafas, la cual se encarga de recibir la información espacial del entorno en una imagen digital. Luego esta imagen se procesa en una CPU para seleccionar la zona de interés; se hace un *zoom* en la zona, se realza el contraste para obtener la información relevante de la escena. A continuación, la información se transforma en una señal de impulsos eléctricos sobre una matriz de 400 electrodos que estimulan las terminales nerviosas de la lengua como reemplazo de la función de la retina. Finalmente, la información llega a la zona de

la función visual cerebral encargada de interpretarla. En la Figura 1 se aprecia el concepto del BrainPort.

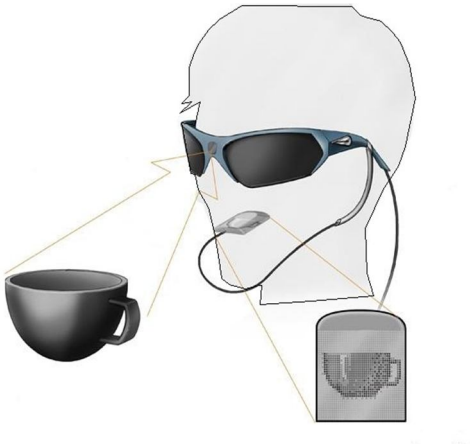

Figura 1. BrainPort[5].

Entre las ventajas que incluye el BrainPort se citan: el incremento en la independencia, mejoras en la seguridad, reconocimiento de objetos y la habilidad para interactuar con el entorno. Algunos pacientes después de mucho tiempo de entrenamiento y uso, han podido leer textos y reconocer pequeños obstáculos en el camino. Entre las desventajas se mencionan el tiempo prolongado de entrenamiento y de aprendizaje para el uso del dispositivo.

2.2 VoiCe

Este dispositivo integrado con un sistema convierte la imagen en patrones de sonido teniendo en cuenta las limitaciones y restricciones del sistema auditivo humano. También se basa en la hipótesis de la sustitución sensorial a través de la plasticidad neuronal del cerebro humano.

El sistema implementa el mapeo de una imagen en sonido preservando la información visual. Para hacerlo, cuenta con una cámara ubicada en unas gafas

[5] http://www.coolplaneta.com/wp-content/uploads/2011/08/brainport-vision-concept-110209-02.jpg

encargadas de recibir la imagen del entorno. Luego, dicha imagen pasa por una etapa de procesamiento que mapea la información visual en patrones de sonidos y finalmente son reproducidos en el individuo. Para mapear las imágenes en sonido, el algoritmo realiza un barrido de la imagen de izquierda a derecha. Por cada columna, se produce un sonido diferente según el nivel de gris y la posición de la fila del píxel. De manera que a un píxel en una altura superior de la imagen le corresponde una frecuencia mayor y la amplitud de la señal depende del nivel de gris del píxel. Finalmente, cada señal de cada píxel se suma, generando en conjunto un sonido que contiene la información visual de la imagen. En la figura 2 se aprecia el concepto del mapeo de la imagen a sonido implementado por VoiCe.

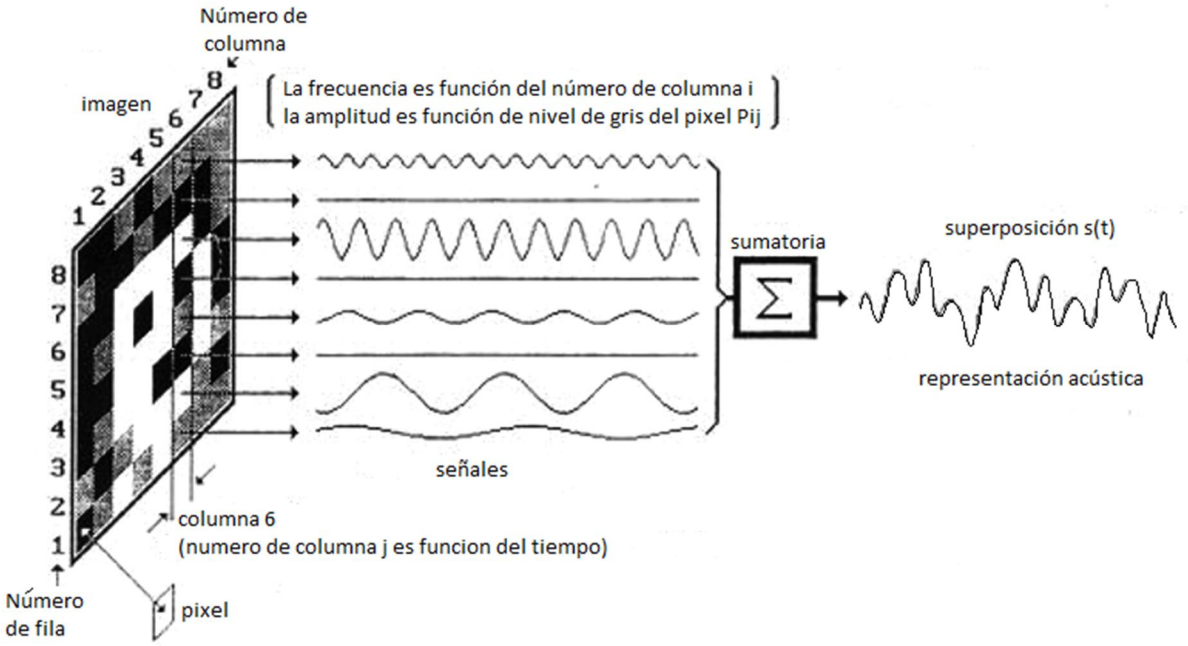

Figura 2. Función de transferencia acústica del VoiCe (Meijer, 1992).

Dentro de los beneficios que ofrece este dispositivo está la cantidad de información que ofrece, el bajo costo económico y el consumo de procesamiento. Sin embargo, requiere de tiempos amplios de aprendizaje, acompañado de entrenadores en orientación y movilidad.

2.3 CASBliP

Cognitive Aid System for Blind People (CASBliP) es un dispositivo, como los demás, diseñado con el propósito de ayudar y asistir a personas en situación de discapacidad visual. El sistema incorporado transforma las imágenes del entorno en mapas acústicos virtuales.

El dispositivo cuenta con un par de cámaras y un sensor de profundidad CMOS-3D para obtener los mapas de profundidad del entorno con algoritmos de estereoscopia y tiempo de vuelo láser (*ToF)*, respectivamente. En una unidad de procesamiento se transforma el mapa de profundidad en un mapa acústico y con base en algoritmos de procesamiento de imágenes se hace el reconocimiento de objetos en movimiento y de personas. Finalmente, el mapa acústico es reproducido en pequeños impulsos de alta frecuencia a manera de "tics" para que la persona pueda configurar una percepción espacial del entorno. Como elementos añadidos al sistema general y con el fin de proporcionar mayor apoyo a las personas con limitaciones visuales, el sistema incluye un GPS y un HPS, para usos en ambientes exteriores. La Figura 3 presenta los componentes del CASBLiP.

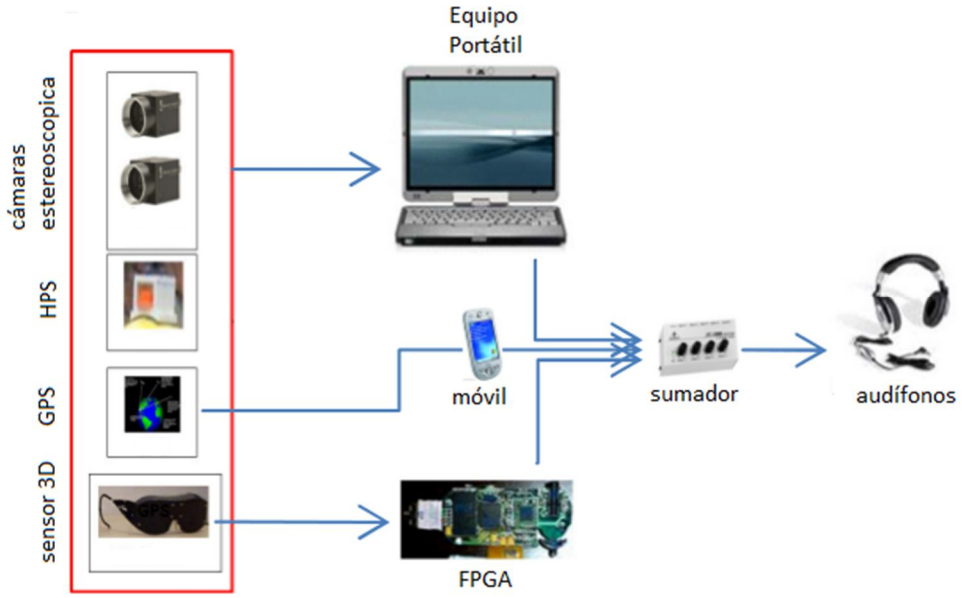

Figura 3. Componentes del CASBliP (Praderas, *et al*, 2009).

Con este sistema la persona puede moverse tanto en espacios abiertos como en espacios cerrados. Detecta la presencia de objetos móviles, permite identificar espacios abiertos (pasillos por donde puedan circular) y objetos estáticos a distancias comprendidas entre 0,5 y 15 metros. Sin embargo el sistema presenta limitaciones: en situaciones estáticas aumenta la dificultad para identificar la posición de más de un objeto presente en la escena. Así mismo, surge una mayor dificultad para que el usuario interprete la información cuando aumenta el número de personas y de objetos en movimiento (Peris-Fajarnés, 2009).

2.4. NAVI

El *Mobile Navigational Aid for Visually Impaired* (NAVI) es un dispositivo que presenta una integración en los modos de navegación en términos micro y macro. El sistema consta de un sensor de profundidad Kinect, un equipo de procesamiento, un actuador vibrotáctil y audífonos. Funciona en dos modos complementarios que se detallan a continuación.

En el primer modo de micro-navegación, se recibe la información del entorno por medio del sensor de profundidad Kinect, luego se procesa en un equipo portátil que determina el camino libre por el que la persona puede transitar, y finalmente, se envía a un actuador vibrotáctil que permite establecer por variaciones de frecuencia la cercanía del obstáculo; a mayor frecuencia de vibración del actuador el objeto está más cerca de la persona.

En el segundo modo, de macro navegación, la información del entorno se capta por medio de la cámara RGB Kinect, mediante etiquetas, *tags* o marcadores de realidad aumentada que contienen información descriptiva del espacio y los objetos. Por ejemplo, las etiquetas con la información de la ubicación de una puerta, un mueble, una oficina, entre otras locaciones. Luego, la información se interpreta en la unidad de procesamiento y finalmente se envía al usuario por medio de sonidos

vocales sintetizados describiendo la zona y su ubicación. Ejemplos de sonidos reproducidos por el sistema son "*Door in 3*", "*2*", "*1*", "*Open door*". La figura 4 se ilustra los dos modos de navegación del que dispone el sistema.

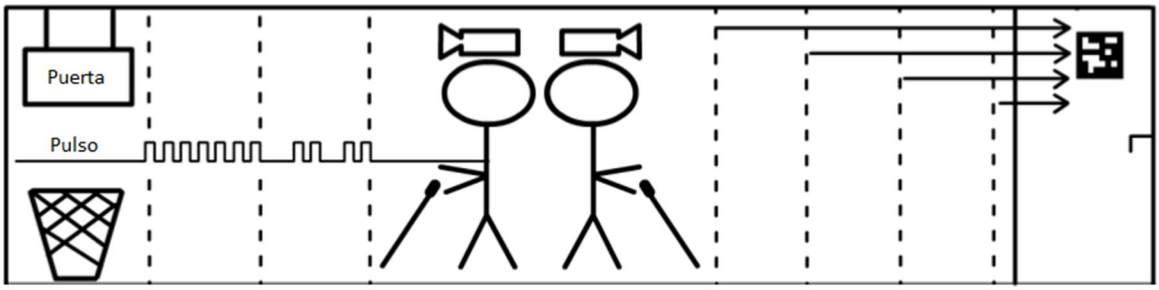

Figura 4. Modos de Navegación del NAVI. (izq) Modo de micro-navegación y (der) Modo de macro-navegación (Zöllner, *et al*, 2011).

Los dispositivos de tiflotecnología anteriormente mencionados asisten desde diferentes enfoques a las personas en situación de discapacidad visual y son el resultado de procesos de investigación y desarrollo. En ese mismo sentido, los resultados que aquí se presentan son el resultado de una investigación que aprovecha los aportes y experiencia en el área de procesamiento de imágenes y aplica conceptos de sustitución sensorial con el propósito de brindar apoyos alternativos para la asistencia a personas con limitaciones visuales. Para esto, se hizo necesario estudiar la configuración de un EAV que permita a las personas identificar la posición de más de un objeto en un entorno dado, a partir de imágenes de mapas de profundidad.

3. ESPACIO ACÚSTICO VIRTUAL

El Espacio Acústico Virtual (EAV) está constituido por un conjunto de sonidos generados por un sistema de manera que quien lo escucha, tiene la sensación de que los objetos y superficies emiten sonidos, ver Figura 5. Según el grupo de investigación y desarrollo en percepción del espacio usando sonidos con aplicación específica para personas ciegas, por medio de un EAV se valida la hipótesis de que una persona pueda tener una experiencia de presencia global del campo de percepción, en cuanto a la forma, las dimensiones y la ubicación de los objetos, es decir, que una persona puede orientarse, ubicar objetos y superficies en el entorno (Rodriguez-Ramos, *et al*, 2006) (Grupo de Investigaciones y Desarrollo en percepción del espacio usando sonidos con aplicación para personas ciegas, 2002).

Para la generación de un EAV es importante considerar la mayor cantidad posible de propiedades y fenómenos presentes en un sonido; entre estos, la espacialización sonora (Gelfand, 2010) similar a la escucha binaural. Gracias a este fenómeno es posible explicar cómo las personas posicionan una fuente de sonido en el espacio. Adicionalmente, para virtualizar una fuente de sonido, se acude a las Funciones de Transferencia *Head Related Transfer Function* (HRTF) (Cheng & Wakefield, 2001), las cuales reciben la posición cartesiana y entregan un sonido relativo a esa posición.

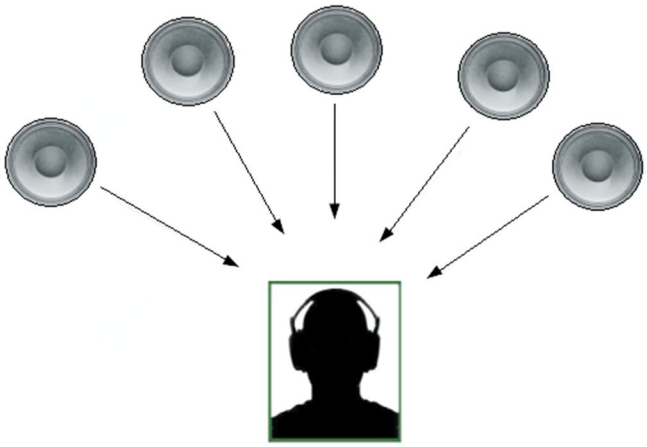

Figura 5. Concepto del Espacio Acústico Virtual.

3.1 Psicoacústica

Para la generación de sonidos a escuchar en el EAV, es adecuado partir de la psicoacústica, puesto que es la rama de la ciencia encargada de estudiar los procesos físicos y psicológicos que ocurren en la percepción de los sonidos en el ser humano. Entre los conceptos que se tuvieron en cuenta para el desarrollo de este proyecto se señalan los límites de percepción en frecuencias y amplitudes, a partir del diagrama de intensidad sonora en función de la frecuencia del sonido (ver Figura 6).

Así mismo, el sistema de audición humano es importante en este trabajo, ya que se le asigna la sustitución sensorial de la función visual. A partir del comportamiento del oído humano fue necesario tener en cuenta algunas limitaciones, restricciones y precauciones durante el diseño, implementación y uso del sistema prototipo desarrollado en esta investigación; ya que en términos generales se puede decir que el sistema de audición humano es un analizador de sonidos o espectrómetro, y el proceso de interpretación sonora ocurre en el cerebro humano.

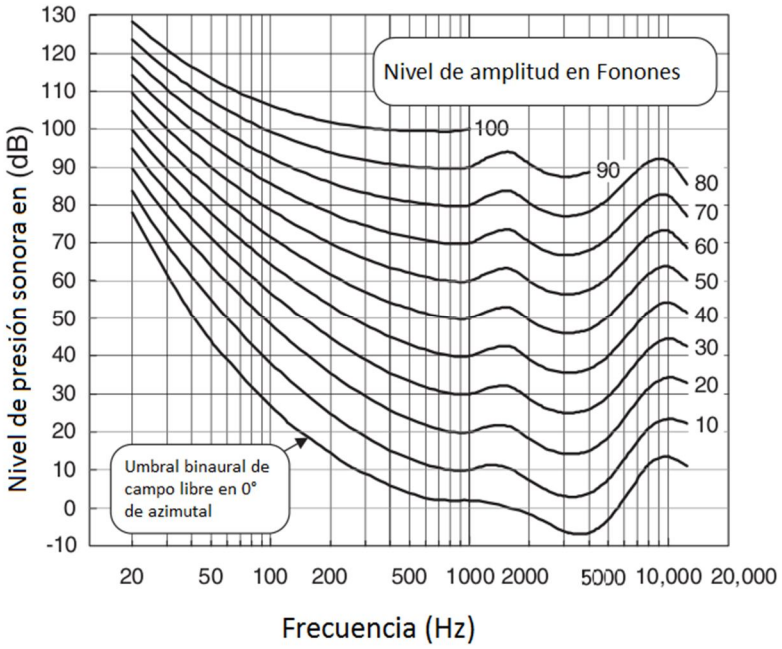

Figura 6. Diagrama de Ganancia del Sistema de Audición Humano (Gelfand, 2010).

El rango humano de audición en amplitud se sitúa entre 0 y 140 dB, y en frecuencia entre 20 Hz y 20 KHz, en promedio. Estos rangos varían de una a otra persona, normalmente la sensibilidad audible disminuye con la edad y la exposición a sonidos de elevada intensidad. En la Figura 6 se puede ver que la zona de audición humana está limitada encima por el umbral de dolor y riesgos de daños, y por abajo con el umbral de audibilidad. Ambos umbrales dependen de la frecuencia de la señal sonora. La mayor sensibilidad del oído humano es para sonidos entre 2 – 5 KHz. Sin embargo, personas de mayor edad pierden sensibilidad a sonidos de alta frecuencia. Por tanto, los sonidos reproducidos por el sistema deben estar en zonas audibles y evitar los posibles daños por los tiempos prolongados de exposición.

3.2 Espacialización sonora.

La espacialización sonora es el proceso mediante el cual se describe como una persona ubica un objeto en su entorno, siendo el cerebro humano el encargado de recibir la señal de ambos oídos y por medio de un análisis diferencial de señales realiza el proceso de localización de fuentes sonoras. Este análisis de diferencias interaurales considera el tiempo y la intensidad de las señales acústicas.

Las diferencias de tiempo interaural (ITD), en inglés, surgen de la diferencia en el tiempo que tarda en llegar la señal acústica a cada oído, ver Figura 7. Las diferencias de intensidad interaural (IID), en inglés, se generan por los efectos de absorción y difracción del sonido debido al tamaño y forma de la cabeza humana. Las ITD y IID operan de forma complementaria en los rangos de frecuencia, para frecuencias inferiores a 1.5 KHz predominan las ITD, y para frecuencias superiores a 1.5 KHz las IID. Por tanto, el rango de frecuencias en el que mejor se aprecian los efectos de ITD y IID para la ubicación de objetos es alrededor de los 1.5 KHz.

La importancia de la espacialización sonora está centrada en brindar sonidos con las diferencias binaurales de tiempo y amplitud apropiadas para que una persona pueda ubicar en su entorno las fuentes sonoras reproducidas por el EAV.

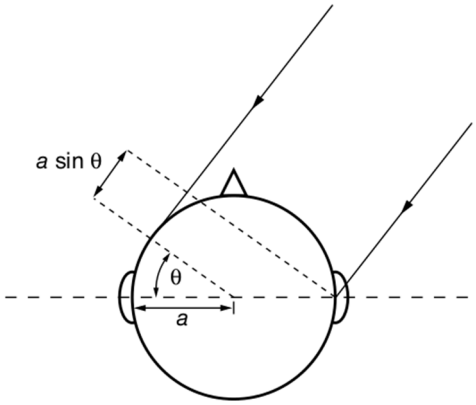

Figura 7. Espacialización sonora. Las ondas de sonido llegan en diferente tiempo e intensidad a ambos oídos [6].

3.3 Funciones de transferencia HRTF

Las funciones de transferencia *Head-Related Transfer Function* (HRTF) o en el dominio temporal *Head-Related Impulse Response* (HRIR), permiten describir la función de transferencia entre cada oído y la fuente de sonido, según su posición en el espacio. Estas funciones tienen en cuenta la forma de la cabeza, el cuerpo y del sistema de audición y por ello varían de una persona a otra. Sin embargo, hay establecidos parámetros promedios que permiten efectos de virtualización muy similares para la mayoría de personas. En la Figura 8 se observa las diferentes señales percibidas en los oídos de las HRIR y HRTF.

[6] http://diamonddissertation.blogspot.com/2010_05_01_archive.html

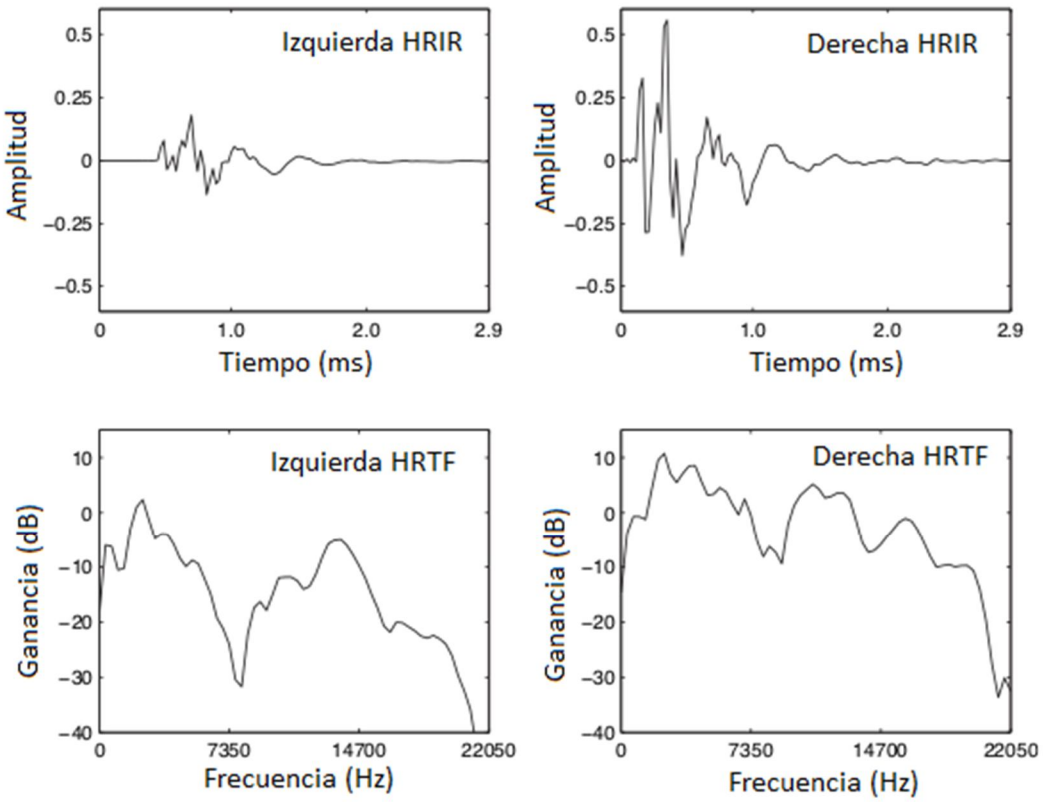

Figura 8. Funciones HRIR y HRTF en oídos izquierdo y derecho para una fuente de sonido ubicada a 40° azimutal derecha y 0° elevación (Cheng & Wakefield, 2001).

Para generar una fuente de sonido virtual que pueda ubicarse en función del EAV, es necesario un generador de señales en el rango audible de frecuencias y amplitudes con la media en ~1.5 KHz y ~60dB; y las funciones de transferencia HRTF o HRIR a partir de la posición, para el caso cartesiana, de los objetos en el entorno. Por tanto, el EAV recibe las posiciones de los objetos y asigna y reproduce los sonidos adecuados para cada fuente de sonido virtual.

4. LA VISION POR COMPUTADOR

Un sistema de visión por computador en este caso, demanda la captura imágenes del entorno en forma de mapas de profundidad, los procesa y analiza para obtener la ubicación de los objetos presentes en el entorno y luego se articulan con el Espacio Acústico Virtual para la reproducción de las fuentes de sonido virtuales. En el desarrollo de la presente investigación se implementó un sistema de visión por computador que consta de tres etapas: 1) Adquisición, 2) Procesamiento y 3) Análisis de los mapas de profundidad, que entregan como resultado las características de los objetos presentes, entre ellas principalmente su ubicación.

4.1 Mapas de profundidad

Un mapa de profundidad es una imagen que describe en una escala de Niveles de Gris (NG), o de color, la ubicación espacial de objetos o superficies de un entorno. De este modo, indica que tan cerca o lejos se encuentran las superficies respecto al dispositivo óptico, es decir, un píxel con NG alto (blanco) corresponde a distancias lejanas; y un NG bajo (gris oscuro) corresponde a distancias cortas, mientras que el Nivel de Negro (cero) corresponde a distancias no resueltas por el dispositivo. En la Figura 9 se aprecia como el computador está más cerca que la bicicleta.

Un mapa de profundidad permite de esta manera conocer la ubicación cartesiana de cada píxel respecto al dispositivo óptico y con ello, la ubicación de los objetos en el entorno.

Figura 9. Mapa de profundidad (izq), Imagen correspondiente al mapa de profundidad (der)

4.2 Obtención de los mapas de profundidad.

Existen diversas técnicas para la obtención de los mapas de profundidad como la triangulación láser y la luz estructurada (Lanman & Taubin, 2009). Esta última consiste en la proyección de un patrón de iluminación estructurada sobre la superficie a capturar, se analiza la deformación del patrón en la superficie y se obtiene el mapa de profundidad, como se muestra en la Figura 10.

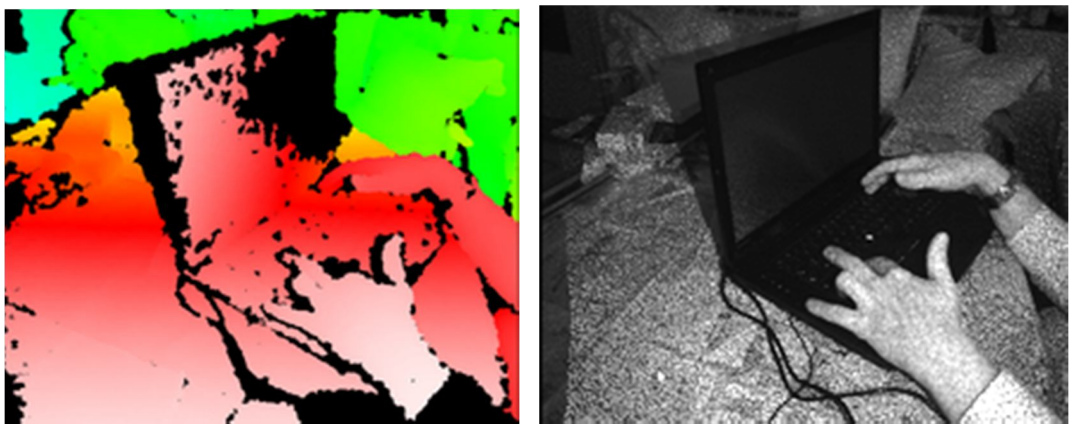

Figura 10. Mapa de profundidad (izq) y Patrón de puntos proyectados (der)[7]

[7] http://en.wikipedia.org/wiki/Kinect

El sistema óptico del dispositivo de captura puede describirse con el modelo pin-hole (Escalera, 2001), ver Figura 11. Este modelo reduce la óptica a un punto situado a la distancia focal de la imagen; por esto, de todos los rayos luminosos que refleja un punto perteneciente a un objeto, solamente es importante el que pasa directamente por la distancia focal. El modelo pin-hole supone que todos los puntos están bien enfocados. Por semejanza de triángulos se puede obtener la relación entre un punto en el espacio y su proyección en la imagen.

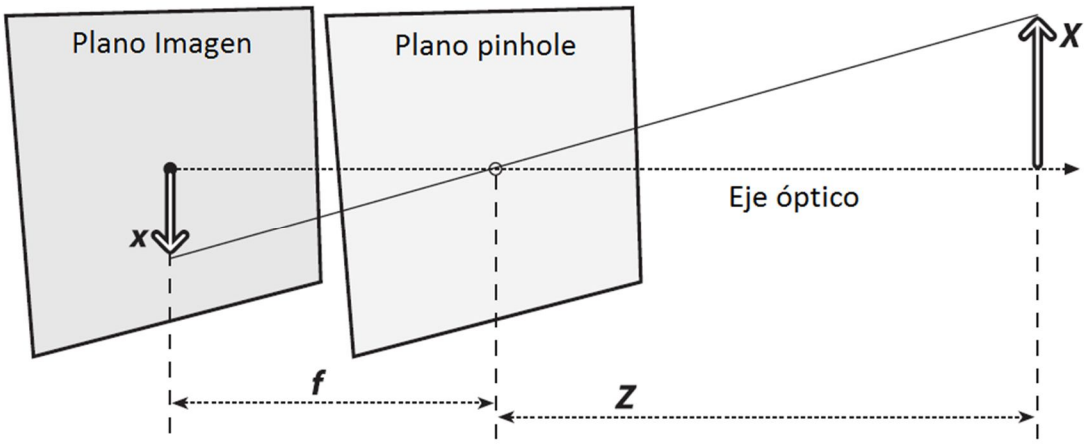

Figura 11. Modelo Pin-Hole (Bradski & Kaehler, 2008)

A partir del mapa de profundidad se puede obtener la nube de puntos que es representada en el espacio. Por medio de la matriz de parámetros intrínsecos A del sistema óptico, se pasan los puntos del espacio imagen al espacio cartesiano.

$$A = \begin{pmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (1)$$

dónde f_x y f_y corresponden a las distancias focales del sistema óptico en píxeles, y u_0 y v_0 corresponden al punto central del plano imagen.

Una vez se obtiene el mapa de profundidad como imagen en NG y se tiene la matriz de los parámetros intrínsecos, se puede procesar y analizar para extraer la

(2)

ubicación espacial de los objetos del entorno, aplicando la inversa de A. De manera que:

$$x = (u - u_0)\frac{z}{f_x}$$
$$y = (v - v_0)\frac{z}{f_y} \quad (3)$$

4.3 Procesamiento y análisis de los mapas de profundidad.

Existen múltiples técnicas de análisis y procesamiento digital de las imágenes que pueden ser aplicados a los mapas de profundidad. Puesto que el propósito del sistema de visión por computador es obtener la posición relativa de los objetos en el entorno, se requiere para esto de un análisis y procesamiento por subetapas, tales como: 1) tratamiento del histograma, 2) segmentación de imágenes, 3) etiquetado de regiones y 4) descripción de objetos.

4.3.1 Tratamiento del Histograma

El histograma de un mapa de profundidad representa la frecuencia y distribución de los píxeles que tienen la misma distancia o NG respecto al dispositivo óptico. En la Figura 12 se aprecia el histograma de una imagen en niveles de gris que está compuesta por objetos (letras) ubicados a diferentes distancias respecto al dispositivo óptico, es decir, las letras A, Z y O están más cerca del dispositivo óptico que las letras B, G, W y E. Mientras que en el Histograma, las letras A, Z y O corresponden a los primeros tres picos de izquierda a derecha, y las letras B, G, W y E a los picos restantes.

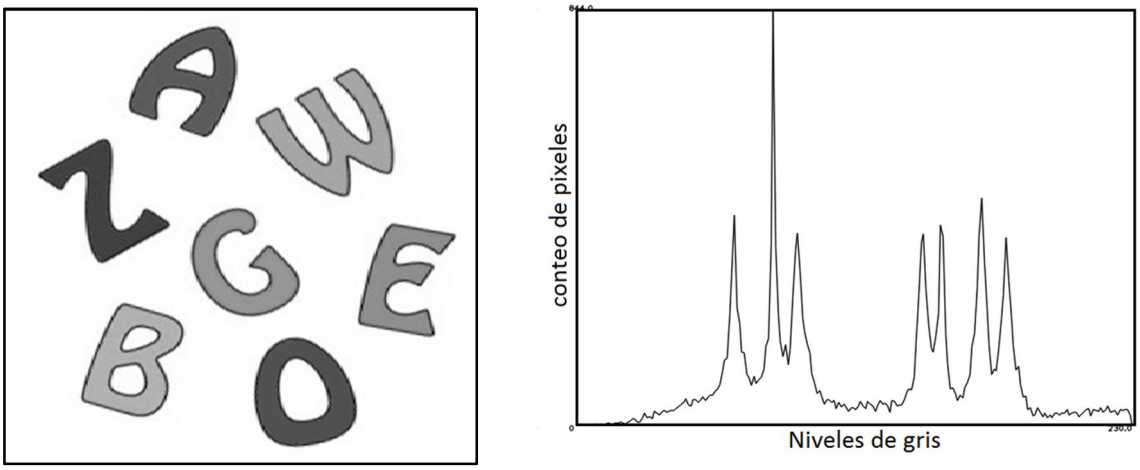

Figura 12. Histograma de una imagen en niveles de gris[8]

A partir de la información entregada por el histograma es factible observar la presencia de objetos en la imagen, puesto que la superficie de un objeto tendrá distancias cercanas en sus vecindades y estas agrupaciones de puntos en el espacio corresponden a agrupaciones de píxeles en el mapa de profundidad. Así, en el histograma se observarán curvas de distribución relacionadas a estas superficies de acuerdo a la distancia en la que se encuentren. De manera que el primer pico corresponde al objeto más cercano y el ultimo pico al objeto más lejano.

4.3.2 Segmentación de Imágenes

La segmentación de imágenes se emplea en las aplicaciones de los sistemas de visión por computador en la subetapa del procesamiento de imágenes. La técnica consiste en separar una o varias regiones de interés presentes en la imagen, diferenciándolas entre sí en dos NG (blanco y negro), en forma binaria 1 y 0, respectivamente. La Figura 13 ilustra este procedimiento, en donde la zona de interés corresponde al color blanco.

[8] http://cuidadoinfantil.net/juegos-de-memoria-para-ninos-sopa-de-letras.html

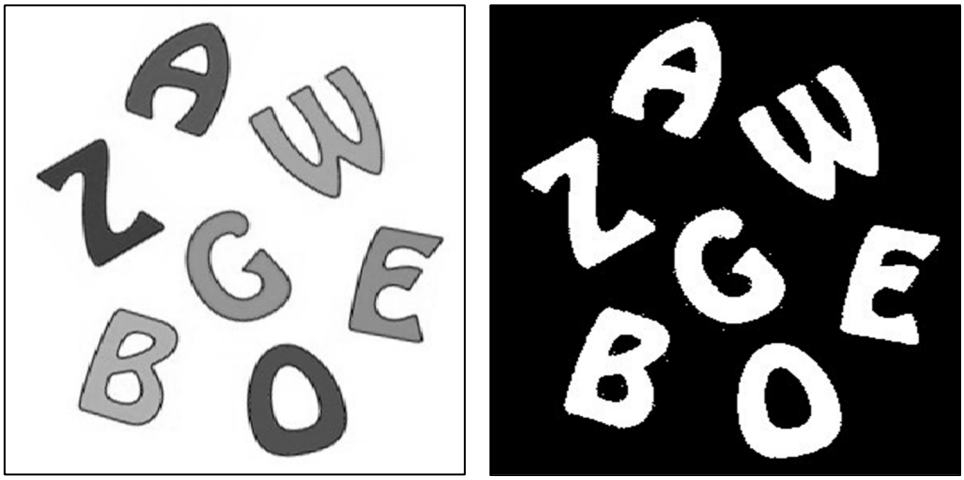

Figura 13. Segmentación de una imagen, separando los objetos del fondo

Para segmentar una imagen es necesario establecer un umbral de segmentación U, los píxeles cuyo NG esté dentro del umbral corresponden a la región de interés, el resto de píxeles se desprecia.

Para determinar el umbral de segmentación y extraer las regiones de los objetos en la escena, se utilizan los umbrales obtenidos del histograma que corresponden a los valores mínimos entre las curvas de distribución. De manera, que los mapas de profundidad quedan separados en distintas imágenes segmentadas, una por cada umbral de segmentación que corresponderá a distancias relacionadas a las posiciones de los objetos. La Figura 13 ilustra la segmentación de las letras respecto al fondo.

4.3.3 Etiquetado de Regiones

Esta técnica permite separar y distinguir las regiones presentes en una misma imagen segmentada teniendo en cuenta los píxeles vecinos y si estos están conectados entre sí, la región conectada corresponde a un objeto.

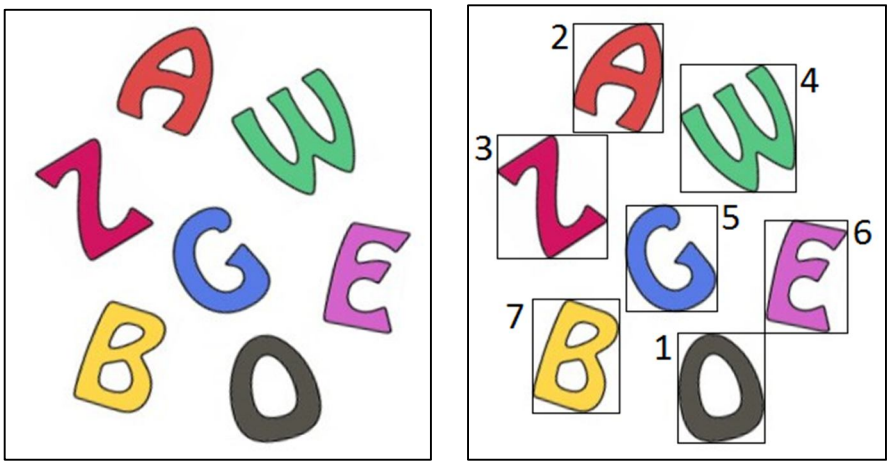

Figura 14. Etiquetado de objetos (letras)

Esta técnica es útil para detectar la presencia de objetos en el mapa de profundidad y diferenciarlos entre sí. En el caso que cada letra de la imagen corresponda a un objeto en el espacio así se deben detectar. En la Figura 14 se aprecia que se puede etiquetar cada letra con un número diferente para distinguirlas entre sí. Además, cada objeto determina un rectángulo delimitador.

4.3.4 Descripción de Objetos: Ubicación Espacial

Para distinguir un objeto de otro, se procede a extraer la información sobre las características de cada región detectada en el mapa de profundidad que corresponde a cada objeto percibido. Las principales características de interés que describen una región son: área, perímetro y el centro geométrico. Para obtener el área, se suman todos los píxeles de la imagen binaria que estén en el interior del objeto etiquetado.

$$area = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} I(i,j)_{internos} \tag{4}$$

El perímetro se obtiene a partir del contorno del objeto mediante la suma de todos los píxeles de la imagen binaria. Estos pixeles se pueden obtener por medio de un análisis local, aplicando el operador gradiente se obtienen las diferencias entre

pixeles en dirección horizontal y vertical, de manera que un cambio de positivo a negativo o viceversa corresponde al borde de una región. Entre los diferentes operadores están el de Roberts, Prewitt, Sobel y Frei-Chen (Escalera, 2001).

$$perímetro = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} I(i,j)_{contorno} \tag{5}$$

El centro geométrico se sitúa a partir de expresiones similares a las de centro de masa de cuerpos planos homogéneos:

$$\bar{\imath} = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \frac{i\, I(i,j)}{A} \tag{6}$$

$$\bar{\jmath} = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \frac{j\, I(i,j)}{A} \tag{7}$$

donde A, es el área del objeto. Con esta información es posible obtener la posición cartesiana del objeto en el entorno.

Las subetapas anteriores permiten obtener la posición de objeto en el plano imagen (i, j, z), que luego al pasar por la matriz de parámetros intrínsecos se pueden aplicar las transformaciones de las ecuaciones 2 y 3 para obtener la posición (x, y, z) del objeto en el entorno. Esta es la información que requiere el EAV para la generación de las fuentes de sonido virtuales.

5. DISEÑO E IMPLEMENTACIÓN DE UN ESPACIO ACÚSTICO VIRTUAL

Para diseñar el EAV a partir de imágenes de mapas de profundidad se debe tener en cuenta consideraciones y requerimientos referentes al rango de frecuencias, amplitudes, funciones de transferencia, dispositivos ópticos y procedimientos de análisis y procesamiento de imágenes mencionados en los capítulos 3 y 4.

5.1 Requerimientos iniciales

Los requerimientos que se tuvieron en cuenta para el diseño del sistema que reproduzca el EAV, fueron los siguientes:

- El rango de frecuencias corresponde al espectro audible, estableciendo una escala con una frecuencia central de 1.5 KHz, para aprovechar las diferencias interaurales de intensidad y tiempo, IID y ITD.
- El rango de amplitudes alrededor de 60 dB con una variación máxima de 20 dB.
- El dispositivo óptico entrega mapas de profundidad entre 0 y 4 metros aproximadamente, lo que permitiría obtener información del entorno cercano frontal, principalmente la posición de los objetos.
- Software para el análisis y procesamiento de imágenes para realizar el tratamiento de los histogramas, la segmentación, el etiquetado y la descripción de objetos, con lo cual se obtiene las posiciones de los objetos en el espacio.
- Software para reproducir sonidos con las funciones de transferencia HRTF y simular la ubicación de objetos en el EAV.
- Además, el sistema debe proporcionar un EAV que permita identificar la ubicación de al menos 2 objetos en el entorno.

A partir de estas consideraciones y requerimientos, se diseñan 3 etapas que conforman el sistema en su totalidad. Las tres etapas corresponden a la

Adquisición de los mapas de profundidad, el Procesamiento y análisis de los mapas de profundidad, y la Virtualización y reproducción del EAV, como se muestra en la Figura 15.

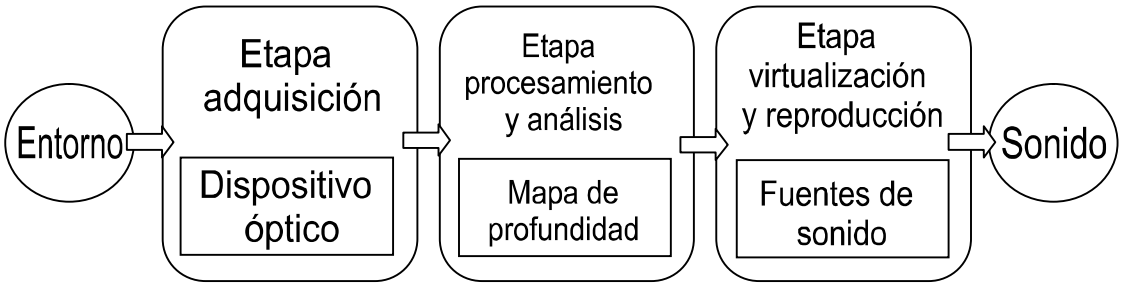

Figura 15. Diagrama del sistema completo que reproduce un EAV a partir de imágenes de mapas de profundidad

5.2 Adquisición de los mapas de profundidad

En la etapa de adquisición se encuentra el sistema óptico que obtiene los mapas de profundidad, tanto la parte hardware como software. Para el desarrollo del proyecto se utilizó el sensor Kinect que es comercializado por Microsoft y desarrollado por PrimeSense. La Figura 16 muestra este dispositivo y un ejemplo del patrón de puntos que suministra; las especificaciones aparecen en la Tabla 4.

5.2.1 Dispositivo Óptico: sensor Kinect

Este sensor utiliza una combinación de métodos para obtener los mapas de profundidad, entre ellos están el principio de la luz estructurada o patrón de iluminación, profundidad por focalización y profundidad por estereoscopía.

Figura 16. Dispositivo Óptico (Kinect) (iqz) [9], Patrón de puntos proyectados por el Kinect (der)[10]

A nivel hardware, el Kinect tiene un proyector de IR que emite un patrón de iluminación con puntos, de manera que las deformaciones de este patrón corresponden a las deformaciones de la superficie, las imágenes son capturadas por una cámara de IR y procesadas internamente por el sensor entregando una imagen de mapa de profundidad. El procesamiento interno para obtener un mapa de profundidad a partir de una imagen IR se realiza de la siguiente manera: El patrón de iluminación permite solucionar problemas de correspondencia entre los puntos de la imagen y puntos en el espacio, por tanto permite utilizar la profundidad por estereoscopía. Posteriormente, se realiza la profundidad por focalización, que analiza cada punto del patrón y según su focalización y densidad de puntos en su vecindad, obtiene la profundidad en la que se encuentra (Freedman, Shpunt, Machline, & Arieli, 2010) .

Tabla 4. Especificaciones técnicas del sensor Kinect

Rango de Distancia	0.5 – 6.0 metros
Resolución Espacial	640x480 píxeles
Resolución temporal	30 fps
Campo de visión	58°
Resolución de profundidad	~ 0.001 metros

[9] http://blog.bricogeek.com/noticias/tecnologia/driver-open-source-espanol-para-microsoft-kinect/
[10] http://wiki.ohm-hochschule.de/roettger/index.php/Projects/KinectRaum-Scanner

Para extraer la imagen de mapa de profundidad del Kinect se requiere de un software especial. Actualmente existen diferentes empresas o grupos que proveen estas herramientas; entre ellas están OpenNI, OpenKinect, Kinect for Windows SDK, entre otros. En este caso se utilizó el Software Development Kit – SDK de OpenNI que es el recomendado por la empresa PrimeSense.

Para enviar el mapa de profundidad de la primera a la segunda etapa es necesario implementar un procedimiento de compatibilidad entre los formatos de imagen en la Etapa de Adquisición (OpenNI) y la Etapa de Procesamiento y Análisis (OpenCV). Para optimizar el sistema en su totalidad se implementaron hilos de programación o *threads*, haciendo que la etapa de adquisición de los mapas de profundidad se realice de forma paralela con las demás etapas.

Finalmente, se obtuvieron los parámetros intrínsecos del Kinect provistos por el grupo de investigaciones *Robot Operating System* (ROS). De manera que la matriz de parámetros intrínsecos tomada de la ecuación 1 empleada para pasar del espacio imagen al espacio cartesiano y viceversa es:

$$A = \begin{pmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 580 & 0 & 320 \\ 0 & 580 & 240 \\ 0 & 0 & 1 \end{pmatrix} \quad (8)$$

donde la distancia focal de la lente IR corresponde a 580 mm/px.

El diagrama de flujo de la etapa de Adquisición de los Mapas de profundidad se representa en la Figura 17.

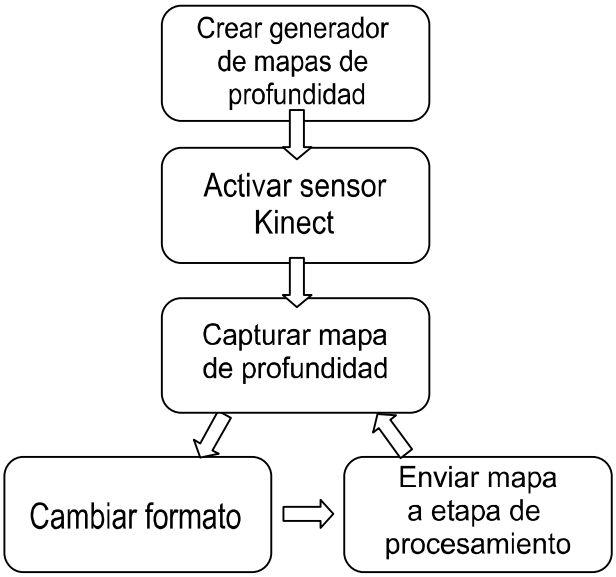

Figura 17. Diagrama de la etapa de adquisición

5.3 Procesamiento y análisis de los mapas de profundidad

Después de obtener los mapas de profundidad de la etapa de adquisición es necesario procesarlos con base en los métodos para el tratamiento del histograma, la segmentación, el etiquetado y la descripción de objetos, con el fin de obtener la posición de los objetos en el entorno, junto con las características necesarias para la Función de Transferencia Acústica (FTA). En esta etapa de procesamiento y análisis que se ilustra en la Figura 18, también se implementa un hilo o *thread* de programación independiente de la etapa de adquisición, con el fin de optimizar los tiempos de ejecución del sistema.

Entre las herramientas software que permiten realizar las subetapas de procesamiento y análisis, se optó por OpenCV. Esta herramienta es una Biblioteca de funciones de programación para realizar visión por computador en tiempo real. Tiene con más de 2500 algoritmos optimizados, licencia de software BSD, con permisos académicos y comerciales. Tiene un grupo estable de soporte y mantenimiento, además de documentación en tutoriales y libros disponibles en internet.

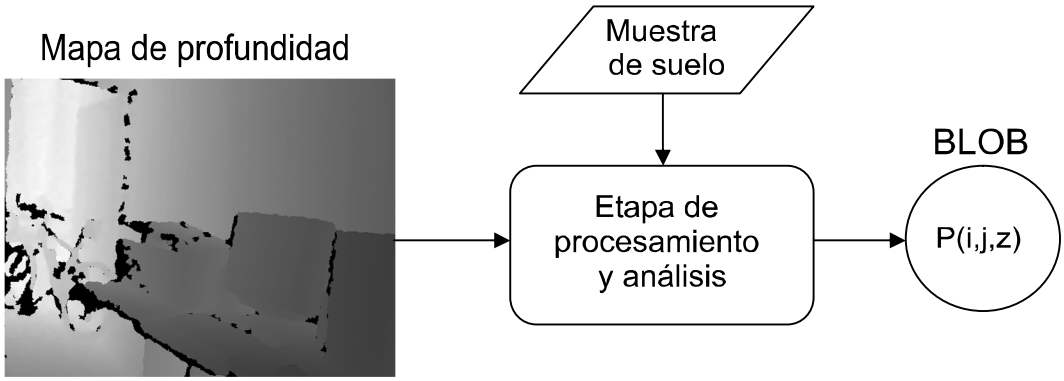

Figura 18. Esquema de la etapa de procesamiento y análisis de los mapas de profundidad

5.3.1 Tratamiento del histograma

El proceso para el tratamiento del histograma permite obtener los umbrales de segmentación, como se planteó en el capítulo 4. Para ello, se debe obtener el histograma del mapa de profundidad, luego extraer las curvas de distribución que corresponden a objetos o regiones del mapa, y finalmente extraer los puntos mínimos entre las curvas de distribución que serán los umbrales de segmentación.

En primera instancia se obtiene el histograma del mapa de profundidad (Escalera, 2001). Puesto que los valores que entrega el Kinect varían entre los 500 mm y 6000 mm, el histograma se adecua para obtener la distribución de las distancias entre los 0.5 y 4.0m, organizados en 600 estados, es decir, que cada estado corresponde aproximadamente a 5 mm.

Posteriormente se extraen las curvas de distribución. Los puntos máximos de estas curvas de distribución son puntos representativos que se deben identificar. Para esto, se filtra primeramente el histograma y se reduce el ruido mediante filtros de máximos y media. Con el filtro de máximos de la expresión 9, a partir de una ventana de extracción, calcula el máximo dentro de ella y lo asigna al valor central

de la misma. Luego se aplica el filtro de media de la expresión 10, con otra ventana de extracción, calcula el promedio y lo asigna al valor central (Escalera, 2001).

Filtro de máximos

$$\boldsymbol{H(x) = max\big(\ H(x-N)\ , H(x+N)\big)} \tag{9}$$

Filtro de media

$$\boldsymbol{H(x) = \sum_{i=0}^{N} \frac{H(x_i)}{N}} \tag{10}$$

donde N, es el tamaño de la ventana de extracción.

Después de filtrar el histograma hasta obtener una curva suave, se procede a obtener la derivada del histograma con el fin de calcular las máximos por los cambios de pendiente de positivo a negativo y cruce por cero. En la Figura 19 se aprecian tres cruces por cero de la pendiente de positivo a negativo, lo que indica la presencia de tres máximos en el histograma.

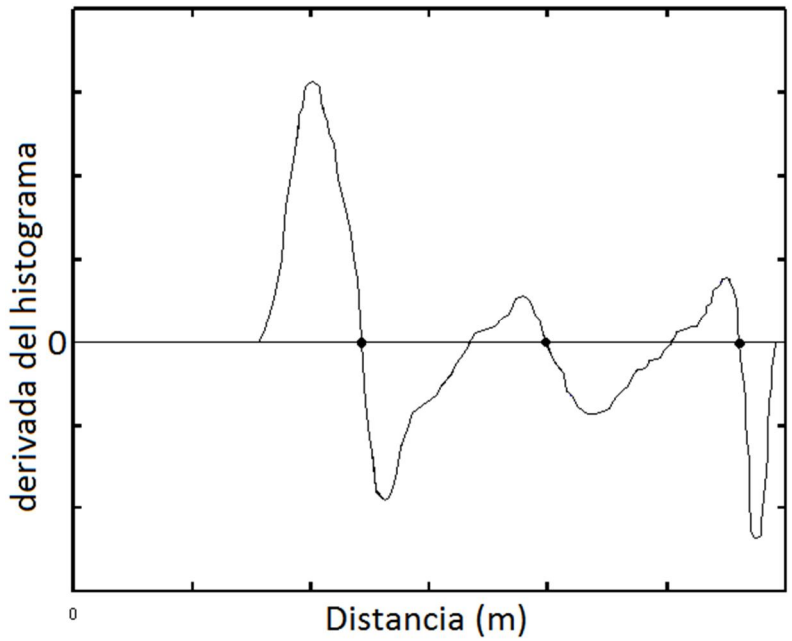

Figura 19. Gráfica del comportamiento de derivada del histograma respecto a la distancia. Los puntos indicados corresponden a máximos en el histograma

Finalmente, se procede a extraer cada mínimo entre dos puntos máximos. El mínimo valor en este rango corresponde al umbral de separación. Los mínimos entre las curvas de distribución corresponden a todos los umbrales de separación. En el ejemplo de la Figura 19, se aprecian dos mínimos o cruces por cero de negativo a positivo, por tanto, dos umbrales de separación. En la Figura 20, se presenta un diagrama que muestra la secuencia del proceso de tratamiento del histograma para la obtención de los umbrales de segmentación.

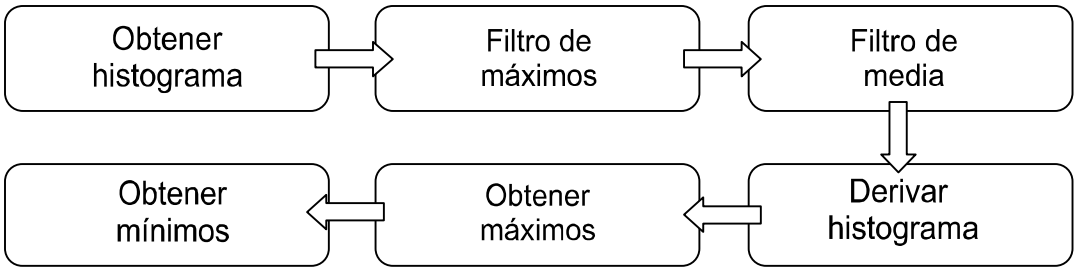

Figura 20. Proceso de Tratamiento del Histograma y obtención de los umbrales de segmentación

5.3.2 Segmentación por ubicación de mínimos.

En la siguiente fase de procesamiento se realiza el segmentado del mapa de profundidad utilizando los umbrales encontrados en el tratamiento del histograma. Así, se obtienen las regiones de los objetos que se encuentran en zonas o distancias similares. El proceso de segmentación binariza el mapa de profundidad, como se mencionó en el capítulo 3. Por cada umbral presente, se obtiene una imagen segmentada.

5.3.3 Segmentación del suelo

Antes de realizar la segmentación de los objetos, es necesario identificar el suelo y segmentarlo. Para ello se requiere de un algoritmo para la detección de suelos o superficies horizontales por las que una persona puede transitar. Esto permite

diferenciar los objetos del suelo. En este caso se utilizó el procedimiento que correlaciona una muestra de suelo $S(u, v)$ con un Mapa de profundidad $I(i, j)$. Si la correlación de un píxel es superior al 80%, dicho píxel corresponde al suelo. En la Figura 21, se observa cómo la muestra de suelo recorre todo el mapa de profundidad verificando si la sección del mapa de profundidad corresponde al suelo.

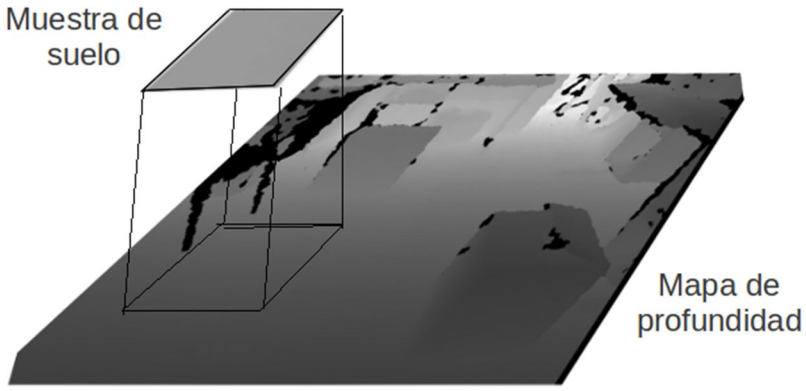

Figura 21. Correlación entre la muestra de suelo (*template*) y el mapa de profundidad

Una vez segmentado el suelo del mapa de profundidad, se procede a segmentar las demás regiones basándose en los umbrales detectados y se verifica que cada región segmentada no corresponda con el suelo. Finalmente, a cada mapa de profundidad segmentado se añade a un vector de mapas segmentados como se ilustra en la Figura 22.

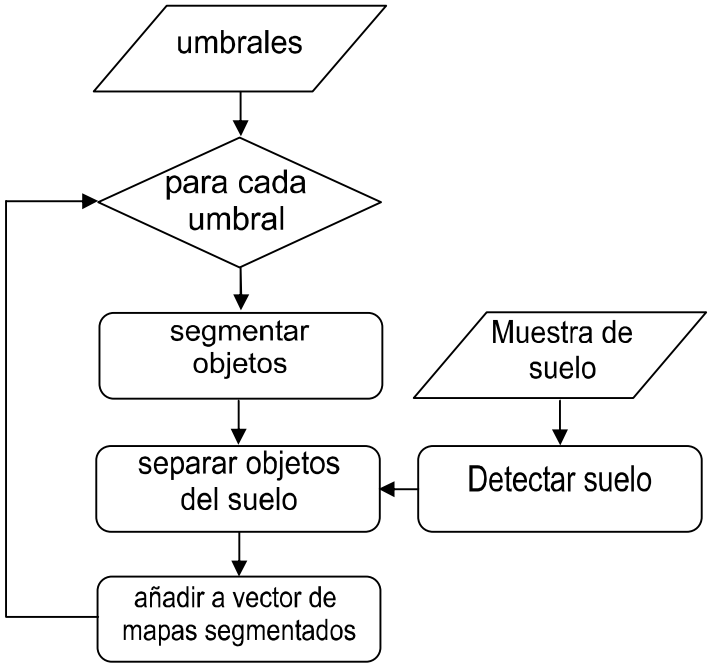

Figura 22. Diagrama de flujo de segmentación de los mapas de profundidad

5.3.4 Etiquetado de regiones por contornos.

El etiquetado de regiones permite separar los objetos entre sí. En cada mapa de profundidad segmentado y almacenado en el vector, se etiquetan los objetos o regiones de superficies allí presentes. Para esto se utiliza un algoritmo de detección de contornos, que obtiene la sucesión de puntos que corresponden a cada región teniendo en cuenta la conectividad entre pixeles. Este algoritmo junto con otros procedimientos permite obtener las características principales de una región (Escalera, 2001).

En el proceso de etiquetado, primero se extraen los contornos del suelo, y luego los contornos de los demás objetos presentes en el mapa de profundidad. Posteriormente son transformados en *Binary Large Objects* (BLOB). Un BLOB es un tipo de dato que almacena el contorno junto con otras características del objeto como son: identificador, área, perímetro, centro geométrico, entre otras, como se ilustra en la Figura 23.

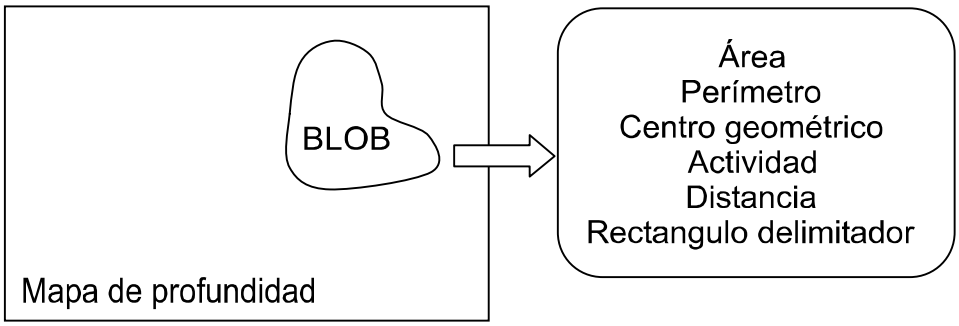

Figura 23. BLOB con sus características principales

5.3.5 Filtrado por área

Después de identificar todos los contornos en el mapa segmentado, es necesario filtrarlos por área para separar los objetos relevantes en la escena y reducir tiempos de procesamiento evitando aquellos que son irrelevantes. En este caso denominamos objeto relevante a un objeto que tenga un área muy grande y que se encuentre a una distancia corta. Dado que uno de los propósitos de este trabajo es que una persona pueda ubicar al menos 2 objetos del entorno, se espera que tales objetos sean los más relevantes. Como criterios para determinar la relevancia se eligieron la distancia y tamaño (área), de manera que relevancia es directamente proporcional al tamaño e inversamente a la distancia.

5.3.6 Ubicación de los objetos y otras características.

Después de obtener los objetos más relevantes según su tamaño, se procede a extraer las principales características que permiten su identificación y seguimiento entre dos mapas de profundidad consecutivos en el tiempo, generando continuidad, estabilidad y robustez temporal en la reproducción del EAV.

Las principales características escogidas en este trabajo son las siguientes:

- Área
- Perímetro - Contorno

- Rectángulo delimitador
- Centro geométrico
- Distancia
- Actividad

Para obtener el área, perímetro y centro geométrico se usan las ecuaciones 1, 2, y 3, en referencia al perímetro se utilizó el detector de bordes de Canny (Escalera, 2001). Para obtener el Rectángulo delimitador, se obtienen los dos puntos diagonales dados por los valores máximos y mínimos tanto en *i* como en *j*. Para obtener la posición del objeto, se hace un promedio de la distancia de los puntos internos P(i,j) del contorno que componen el objeto en el mapa de profundidad. Luego, se extrae la distancia del mapa de profundidad en P(i,j). El promedio de estas distancias se toma como la distancia del sensor al objeto.

Cuando se obtienen las características, nuevamente se filtran los objetos en función del área normalizada para reducir el número de BLOB, eliminando los irrelevantes

$$area\ normalizada = \frac{area}{distancia\ focal^2} \quad (11)$$

Este filtro separa objetos cercanos de gran tamaño de objetos cercanos de tamaño reducido, considerados como irrelevantes. Posteriormente, se realiza un filtrado respecto al parámetro de la distancia, con este se eliminan los objetos situados a distancias superiores a los 4m. Finalmente, los BLOB se ordenan por distancia, del más cercano al más lejano; de manera que el primer BLOB de la lista es el más relevante de todos.

5.3.7 Seguimiento de BLOB

Debido a que ocurren cambios entre dos mapas de profundidad continuos en el tiempo, es necesario que haya una relación entre estos y los BLOB. Por tanto, se comparan las características de los BLOB de un mapa de profundidad anterior con

el siguiente y aquellos que tengan una similitud entre sí superior al 70%, copian su identificación para conservar su permanencia temporal. El criterio que permite una mejor correspondencia entre BLOB es el de la mínima distancia (López, 2011), que supone que un objeto no se moverá tan rápido entre dos mapas de profundidad consecutivos de forma que su nueva ubicación será próxima a la anterior.

A medida que un BLOB permanece en el tiempo aumenta su actividad y cuando esta supera un tiempo mínimo, alrededor de los 200 ms (ó 2 mapas de profundidad a 10 fps), se dice que el BLOB permanece el tiempo suficiente como para no ser causado por un ruido externo. De manera similar, si el BLOB desaparece de escena por un tiempo superior a 300 ms (ó 3 mapas de profundidad a 10 fps), se dice que el BLOB ha perdido su actividad y es inactivo. Este procedimiento permite aumentar la histéresis del sistema y otorga estabilidad espacial y temporal en la detección de objetos presentes en el entorno, ya que pueden existir regiones detectadas por pequeños espacios de tiempo que introducen ruido al sistema y no aportan información estable del entorno.

Para el manejo de BLOB en cuanto al etiquetado, filtrado y seguimiento de regiones en un mapa de profundidad se implementó una biblioteca de funciones, donde cada BLOB cumple con las características mencionadas anteriormente.

Al finalizar la Etapa de Procesamiento y Análisis de los Mapas de Profundidad se obtiene una lista de BLOB ordenada de mayor a menor relevancia. Cada BLOB contiene la información de la ubicación en el plano imagen en coordenadas (*i,j,z*), donde c(*i,j*) es el centro geométrico, y z es el promedio de la distancia sensor-objeto. Con éstas coordenadas se utiliza la matriz de parámetros intrínsecos (ecuación 8) para obtener la ubicación de cada objeto en el plano cartesiano de coordenadas (*x,y,z*). Todo este proceso que se ilustra en la Figura 24 es indispensable para asignar a las fuentes de sonido virtuales los parámetros acústicos y de reproducción necesarios para su reproducción en el EAV.

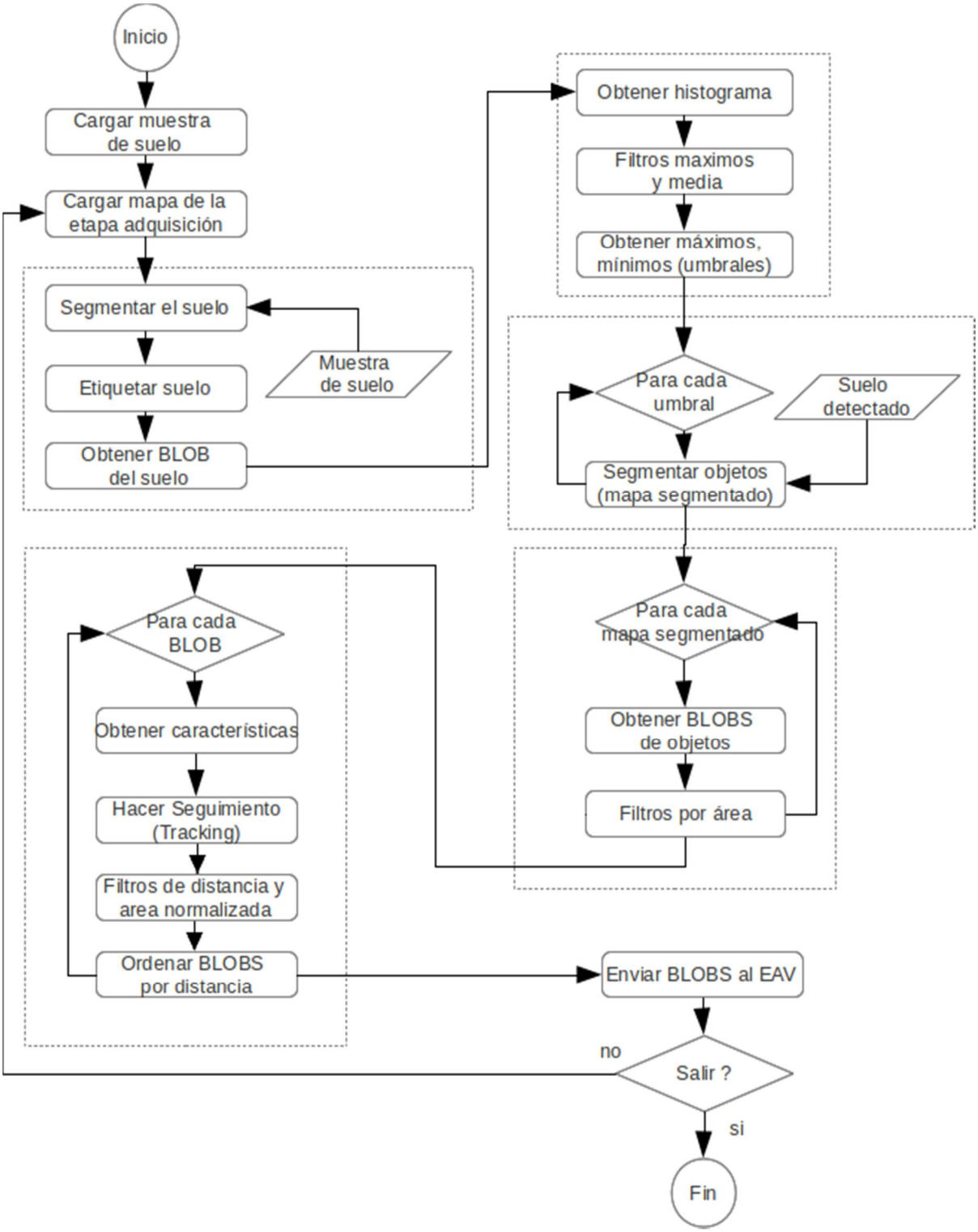

Figura 24. Diagrama de flujo de la etapa de procesamiento y análisis

5.4 Virtualización y reproducción del EAV

La tercera y última etapa del sistema consiste en virtualizar las fuentes de sonido dentro del EAV, para asignar a cada objeto un determinado sonido con base en una Función de Transferencia Acústica (FTA), y reproducirlos para que una persona pueda ubicar 2 o más objetos en el entorno como se muestra en la Figura 25.

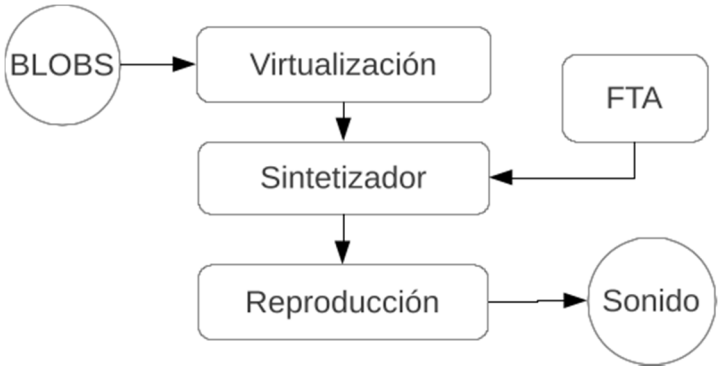

Figura 25. Etapa de virtualización y reproducción

En la implementación de esta etapa de acuerdo con los requerimientos iniciales, se escogió OpenAL como herramienta software que permite la reproducción de sonidos en 3D, por medio del uso de las funciones de transferencia HRTF. OpenAL es una biblioteca con funciones de reproducción de sonidos desarrollada especialmente para juegos; tiene una arquitectura similar a la de OpenGL. Principalmente, reproduce los sonidos a partir de archivos grabados o valores sintetizados, permitiendo hacer *looping* y *streaming*. Además, incluye el manejo de fuentes de sonidos (*sources*), un usuario (*listener*) y un entorno de reproducción (*context*) (Figura 26).

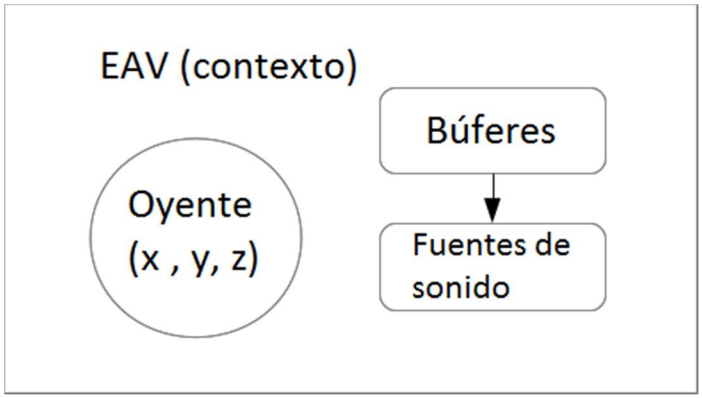

Figura 26. Contexto del OpenAL

5.4.1 Virtualización de las fuentes de sonido

El proceso de virtualización de las fuentes de sonido consiste en distribuirlas y activarlas dentro del EAV, según los BLOB detectados. Para virtualizar las fuentes de sonido, primero deben ser generadas en el EAV (ó contexto de OpenAL) con los parámetros acústicos y de reproducción presentados en la Tabla 5.

Tabla 5. Parámetros de una fuente de sonido virtual

Acústicos	**Reproducción**
Posición	Tamaño del búfer
Señal	Formato
Pulso de modulación	Frecuencia de muestreo
Amplitud	Búferes procesados
Frecuencia	Búferes en cola
Frecuencia de Modulación	
Tiempo	

Cuando se generan las fuentes de sonido, los parámetros de formato, frecuencia de muestreo y tamaño del búfer, permanecen constantes durante la ejecución del sistema. La posición se asigna cuando se virtualiza la fuente de sonido, los demás parámetros acústicos toman valor al pasar por la FTA. Finalmente, los parámetros

de búferes procesados y en cola son asignados y modificados durante la reproducción del sistema.

Después de generar las fuentes de sonido, son distribuidas en una la malla de virtualización en un arreglo de 105 fuentes de sonido distribuidas en una matriz de 15x7, como se aprecia en la Figura 27. En el proceso de virtualización, se recorre la posición de cada fuente sonora y se verifica si pertenece al contorno de cada BLOB detectado. Si una fuente sonora pertenece a un BLOB, se activa y procede a la FTA, de lo contrario se desactiva y no se reproduce ningún sonido. En la Figura 28 se presenta el diagrama de flujo del la subetapa de virtualización.

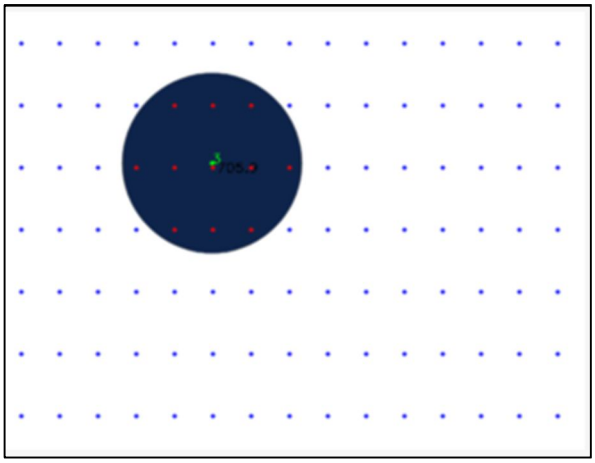

Figura 27. Distribución de las fuentes de sonido, malla de virtualización. Las fuentes activadas están dentro del circulo, el resto están inactivas

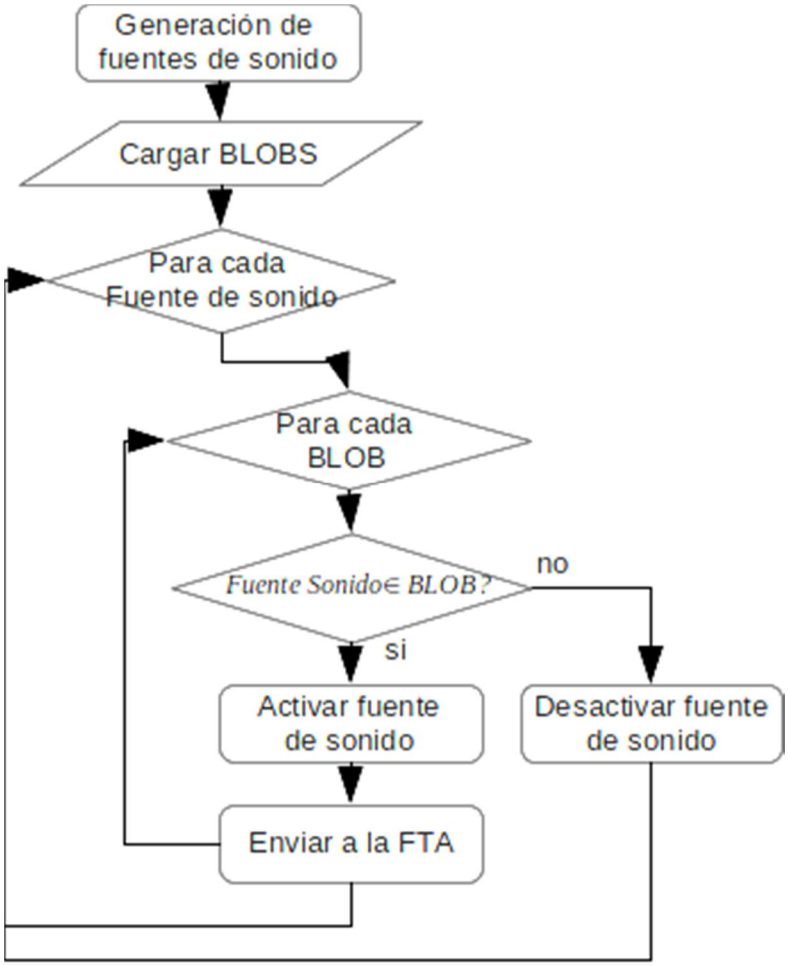

Figura 28. Diagrama de flujo del proceso de virtualización

5.4.2 Función de Transferencia Acústica.

La FTA es un conjunto de funciones que relaciona los parámetros acústicos de las fuentes de sonido con las propiedades de área, posición, distancia, orden de ubicación de los BLOBs para sintetizar el sonido de la fuente. El proceso de sintetización requiere una señal generadora y una Lookup Table (LUT), o Tabla de consulta (Bristow-Johnson, 1996). La función generadora debe tener la propiedad de unicidad a lo largo del periodo de la señal. Por tanto, se escogió la función "diente de sierra" - *Saw Tooth* como la que se ilustra en la Figura 29.

$$s(t) = 2\big(t - floor(t)\big) - 1 \qquad (12)$$

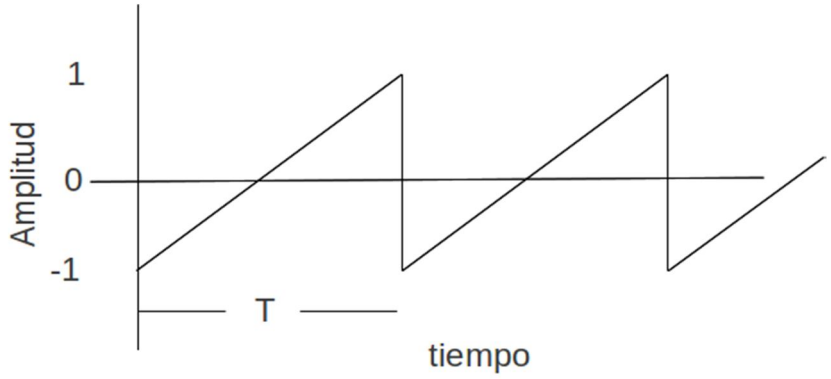

Figura 29. Función generadora para la síntesis del sonido, diente de sierra

A partir, de la señal generadora es posible obtener otras señales con base en la tabla de consulta. Para ello, primero se determinan los Coeficientes de Fourier de una señal, en este caso se usan 15 coeficientes por señal. Luego, por Series de Fourier se llena la LUT para un periodo completo ocupando todos los valores de la misma. El uso de la señal generadora permite modificar la frecuencia de la señal resultante sin necesidad de recalcular la Serie de Fourier para obtener la señal, como se muestra en la Figura 30 y las acciones de la FTA se muestran en la Figura 31.

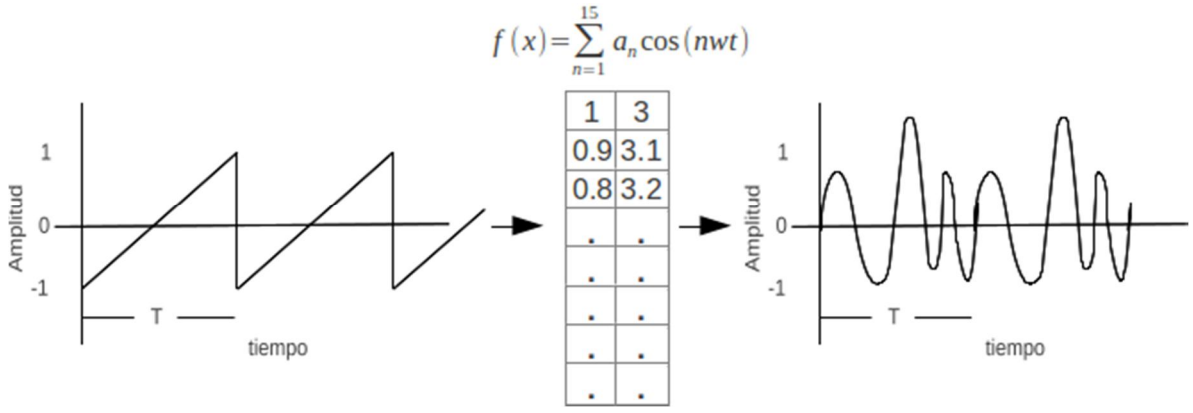

Figura 30. Sintetización del sonido por tablas de consulta (LUT)

La amplitud de la señal es una función que depende de la distancia del BLOB y disminuye con el cuadrado de la distancia.

$$A = F(z) = A_0 \frac{1}{z^2} \quad (13)$$

La frecuencia de la señal es una función que depende de la altura del BLOB en el mapa de profundidad, no respecto al tamaño. Para definir el rango de frecuencias a reproducir, se escogió una escala psicoacústica (Gelfand, 2010), que se basa en la sensación y la psicoacústica del sonido en una persona. La escala propone unas bandas críticas de audición que se tienen en cuenta al momento de escoger las siete (7) frecuencias, que corresponden a las 7 posiciones posibles dentro de la malla de virtualización. La función permite establecer una función entre altura de la posición y la frecuencia; a mayor altura de la posición (*y*), mayor frecuencia. De manera, que a los objetos ubicados en la parte superior del mapa de profundidad les corresponden frecuencias altas y en la parte baja del mapa de profundidad les corresponden frecuencias bajas.

$$w = F(y) = 2\pi f_i \frac{y}{N} \quad (14)$$

donde, fi es la frecuencia i-ésima en la escala de notas definidas al inicio de la ejecución del sistema y puede ser modificada según el usuario y N es el tamaño del mapa de profundidad.

Aunque las señales sean diferentes, se realizan modificaciones en el timbre de las señales con efectos de modulación en amplitud. La frecuencia de modulación en amplitud es inversamente proporcional a la relevancia del BLOB, de manera que las frecuencias de modulación bajas para BLOB muy relevantes. En el caso que dos objetos se encuentren a distancias similares y se aprecie un enmascaramiento por amplitud (Gelfand, 2010), la frecuencia de modulación se hace diferente dependiendo de la relevancia, de esta manera se puede distinguir un objeto de otro. Es necesario que los objetos tengan diferentes frecuencias de modulación, separadas entre sí lo suficiente para que sean perceptibles por el oído humano.

Las frecuencias de modulación se ajustan experimentalmente y están comprendidas entre 1 y 100 Hz.

El enmascaramiento por amplitud puede ser crítico cuando aumenta el número de objetos en el entorno, entonces, el pulso de modulación permite separar los sonidos de cada BLOB entre sí temporalmente. Esto aumenta la posibilidad de distinguir los objetos, diferenciarlos en un periodo de reproducción e identificar cual puede estar más cerca. El efecto es similar a las notas musicales en un pentagrama.

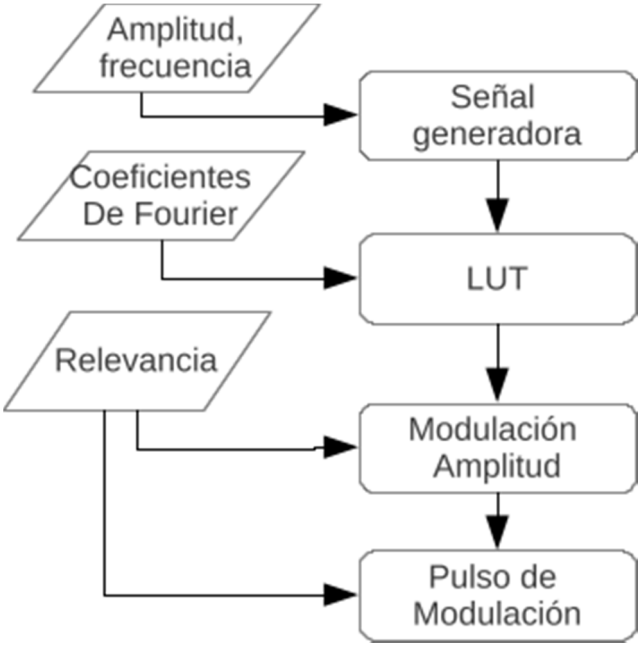

Figura 31. Sintetizador de sonido que implementa las Funciones de Transferencia Acústicas - FTA

5.4.3 Reproducción del Espacio Acústico Virtual

Para reproducir una señal sintetizada se requiere del uso de *streaming* (Wikipedia, Streaming, 2012), que es una forma de reproducción continua temporal empleando el uso de búferes o almacenadores de datos. Mientras se van sintetizando

muestras de sonido, se van almacenando en búferes que son reproducidos con un determinado tiempo de latencia o retardo.

En esta subetapa, la muestra de la señal generada por el sintetizador es almacenada en los búferes de las fuentes de sonido, si al menos la fuente tiene una muestra es reproducida. Al finalizar el ciclo de reproducción, se verifica cuantos búferes han sido reproducidos para luego liberar memoria y habilitar espacio para nuevas muestras de sonido (Figura 32).

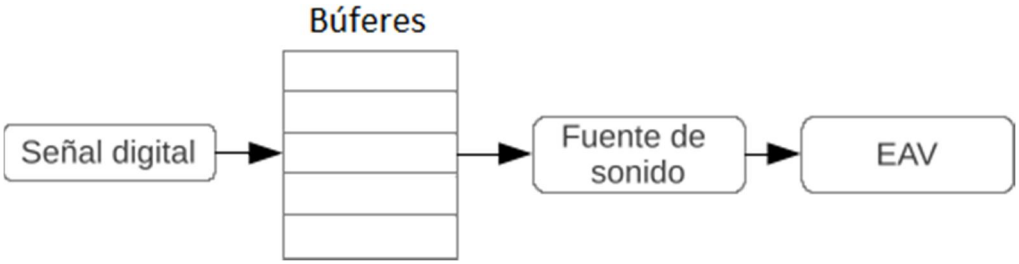

Figura 32. Reproducción del Espacio Acústico Virtual

Finalmente, en cuanto al software implementado, se tiene que el sistema genera un EAV indicando la posición de objetos en el entorno a partir de imágenes de mapas de profundidad. La Figura 33 muestra el diagrama de flujo de la etapa de virtualización y reproducción del EAV.

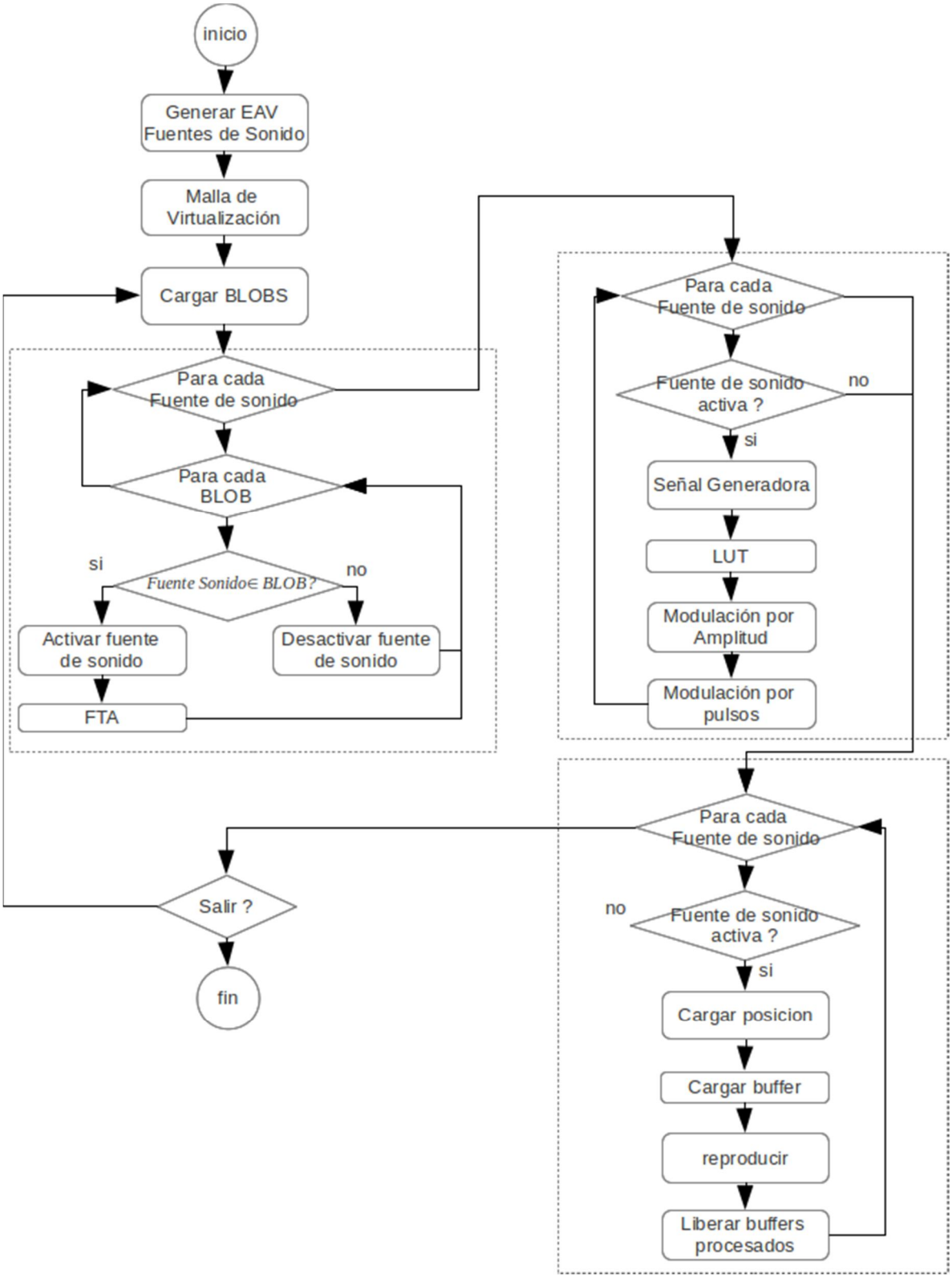

Figura 33. Diagrama de flujo de la etapa de virtualización y reproducción

5.5 Implementación hardware y software del prototipo

En la implementación del prototipo de laboratorio, se utilizaron los componentes presentes en la Tabla 6.

Tabla 6. Componentes del Prototipo Implementado

Componente	Características
Portátil Dell Vostro	Procesador Intel corei3 – 2.3 GHz Memoria RAM: 4GB
Sensor Kinect	640x480@30fps
Audífonos	
Batería	12V - 4AH
Ventilador	12V

Se instalan las dependencias software en un equipo portátil Dell Vostro para el funcionamiento del sistema.

Tabla 7. Lista de herramientas software

Sistema Operativo Ubuntu 11.04 (oneiric)
Compilador g++-4.6
Eclipse Developers for C++ - Juno
QtCreator 4
Driver Kinect ps_engine v5.0.3
OpenCV v2.3.1
OpenAL v1.13.1
OpenNI v1.3.2

El prototipo resultante se muestra en la Figura 34.

Figura 34. Prototipo de laboratorio, reproduce un EAV a partir de imágenes de mapas de profundidad

6. RESULTADOS Y ANÁLISIS

A continuación se presenta un ejemplo del procesamiento en cada una de sus etapas y luego los resultados del sistema tras unas pruebas piloto.

6.1 Adquisición de los mapas de profundidad

En la Figura 35, se aprecia el resultado de la captura de imágenes de mapas de profundidad por medio de la etapa de adquisición.

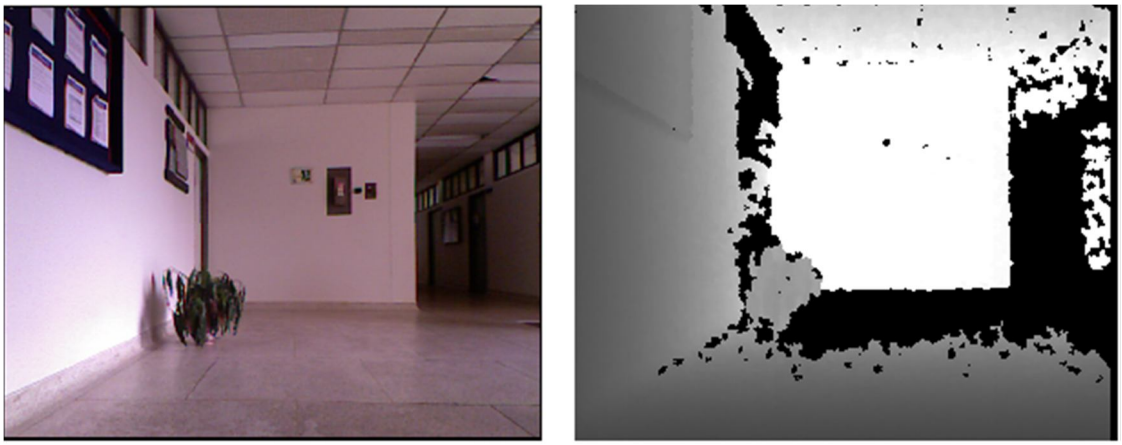

Figura 35. Resultado etapa adquisición

Esta etapa captura imágenes de mapas de profundidad a una tasa de 30 fps, velocidad permitida por el Kinect.

6.2 Procesamiento de los mapas de profundidad

Para la etapa de procesamiento y análisis se muestran imágenes obtenidas correspondientes al tratamiento del histograma, la segmentación del suelo y objetos, el etiquetado de regiones y el seguimiento de regiones.

6.2.1 Tratamiento del histograma

La Figura 36 muestra un mapa de profundidad con su correspondiente histograma. En el mapa de profundidad, imagen (b), se tiene la descripción de un salón con una pared al fondo y una puerta a la izquierda de la imagen, adicionalmente en la imagen (d) se encuentra la presencia de una caja en medio del salón como objeto a detectar. En los histogramas se aprecia la diferencia entre los resultados sin objeto (a) y con objeto (c)., lo que permite su detección y posterior ubicación.

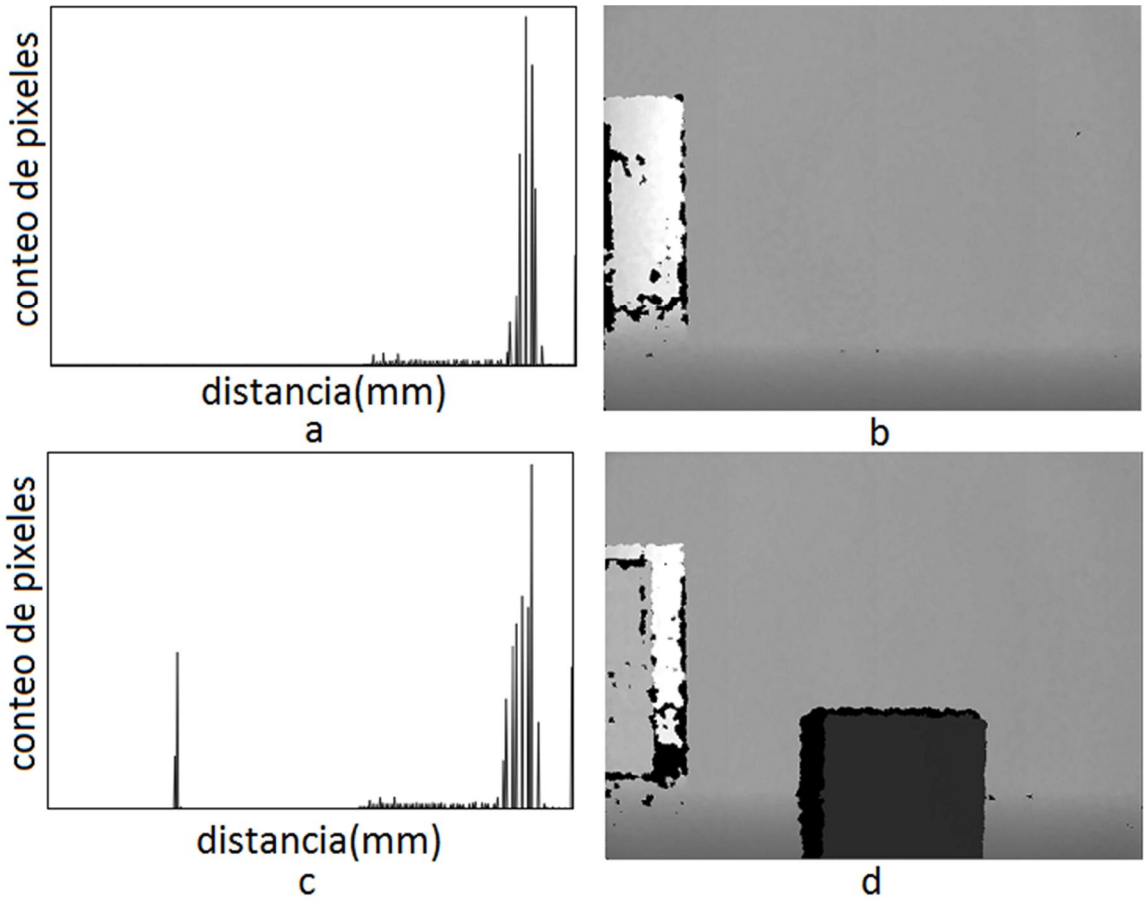

Figura 36. Resultado de la obtención de histogramas (H) de los mapas de profundidad (MP). (a) H sin objeto, (b) MP sin objeto, (c) H con objeto, y (d) MP con objeto.

En la Figura 37 se aprecia el tratamiento que se aplica al histograma obtenido previamente para la extracción de los puntos máximos y mínimos y obtener los umbrales de segmentación. En la imagen (a) se tiene el histograma original, donde el primer pico corresponde al objeto, después se aplica el filtro de máximos (b) que permite amplificar el ancho de banda de ese pico, seguido a esto se realiza el filtrado de media (c) y finalmente la obtención de los puntos máximos (puntos negros) y mínimos, que son los umbrales de segmentación (puntos blancos) (d).

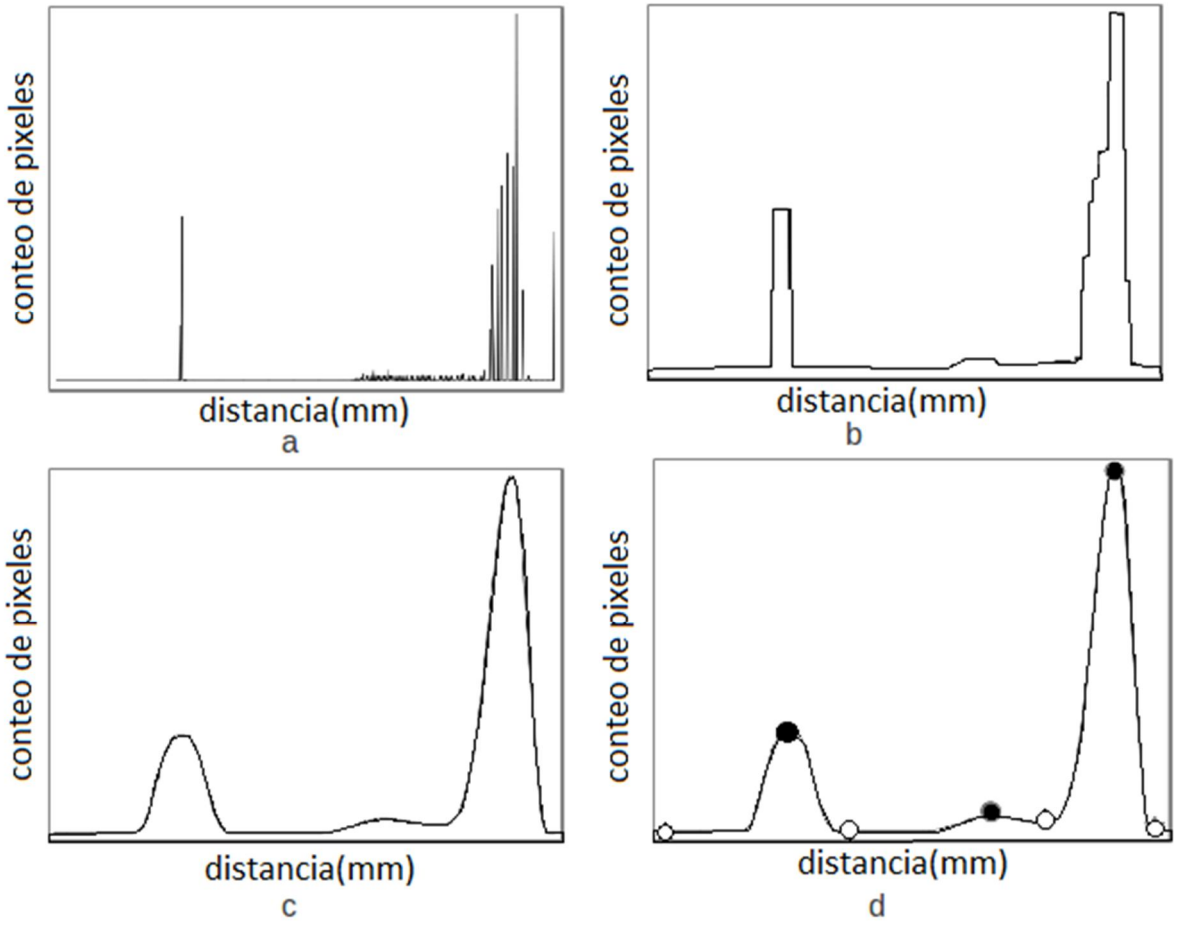

Figura 37. Resultado del tratamiento del histograma. (a) Histograma original, (b) Filtro de máximos, (c) Filtro de media, (d) Obtención de máximos y mínimos

Para este mapa de profundidad y según la cantidad de puntos máximos en el histograma, se tienen tres regiones a segmentar.

6.2.2 Segmentación de regiones

Una vez obtenidos los umbrales de segmentación se procede a la extracción de los mapas de profundidad segmentados como se aprecia en la Figura 38. De acuerdo al tratamiento del histograma anterior se obtienen tres mapas de profundidad segmentados que corresponden al suelo (a), al objeto (b) y a la pared (c). La puerta no es detectada puesto que se encuentra a una distancia superior a los 4m de distancia.

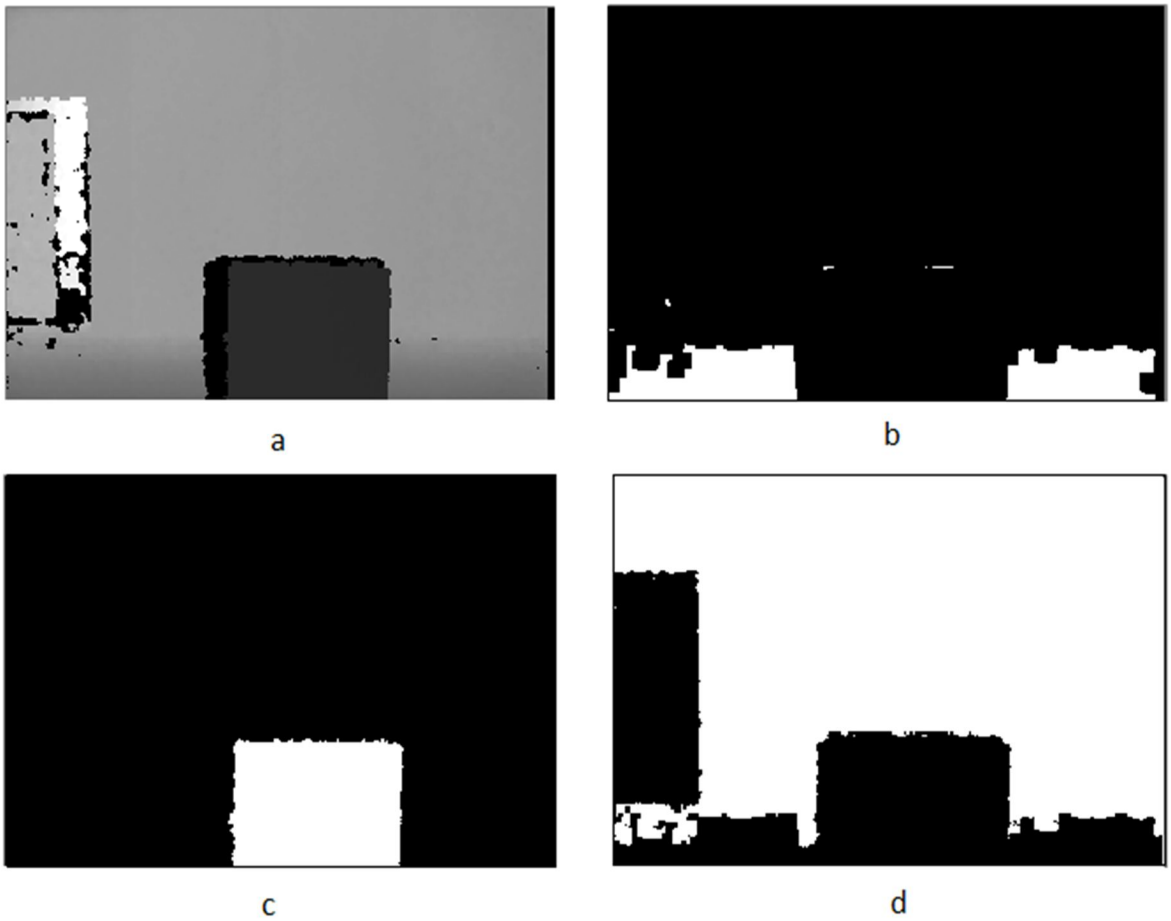

Figura 38. Resultado de la segmentación de suelo y regiones. (a) Mapa de profundidad, (b) Segmentación del suelo, (c) Segmentación región 1 (caja), (d) Segmentación región 2 (pared)

6.2.3 Etiquetado de regiones y seguimiento

Después de la segmentación se procede al etiquetado de regiones. En la Figura 39 se aprecia como un objeto (caja) en (a) es segmentado y sus regiones son detectadas en (b), donde a cada región es asignado un color diferente para tener una apreciación visual del etiquetado de regiones. En cuanto al seguimiento se aprecia como el objeto de (a) ubicado en la posición 1, se traslada hacia la izquierda a una posición 2 en (c) y se observa como en (d) se conserva su detección, identificación y color de etiquetado durante su desplazamiento.

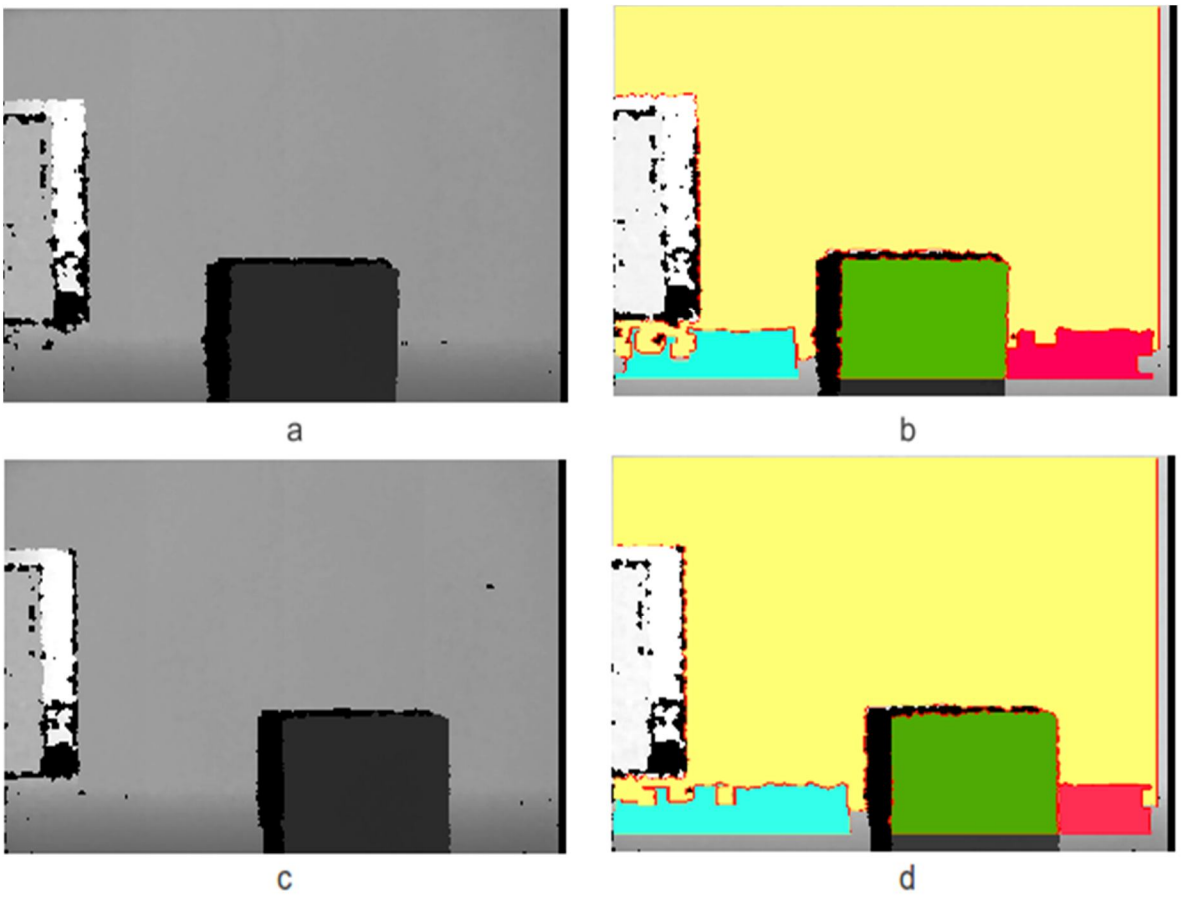

Figura 39. Resultado del etiquetado y seguimiento de regiones. (a) Mapa de profundidad anterior, (b) Etiquetado de regiones, (c) Mapa profundidad posterior, y (d) seguimiento de regiones.

6.2.4 Extracción de características

Después de etiquetar las regiones es necesario extraer las características de los BLOB. La Tabla 8 consigna el resultado de las características de las regiones presentes en la Figura 39. El BLOB con identificación (Id) 1 de mayor área y distancia corresponde a la pared, el BLOB con id 2 corresponde al objeto y los BLOBS 3 y 4 corresponden al suelo.

Tabla 8. Resultados de las características de los BLOB

Id	Área(px)	Perímetro (px)	Rectángulo delimitador (x,y,w,h)	Centro Geométrico (x,y)	Distancia (mm)
1	205083	2908	(1, 1, 619, 431)	(329,181)	3510
2	27532	748	(347, 307, 194, 153)	(412, 386)	1719
3	13538	818	(1, 397, 273, 63)	(145, 432)	2528
4	5915	350	(507, 398, 106, 62)	(557, 429)	2966

6.2.5 Filtros de BLOB por distancia y área normalizada

Posteriormente, los BLOB son filtrados en una distancia entre 300 y 4000 mm, distancia en la que se encuentran los objetos más relevantes, además son filtrados por el parámetro de área normalizada entre 0.04 y 1.0, para eliminar objetos muy pequeños, y finalmente se ordenan por relevancia. El resultado del proceso de filtrado de los BLOB de la Tabla 8 se muestra en la Tabla 9, donde los identificadores han cambiado y el BLOB 1 corresponde al objeto, el 2 a la pared y el 3 al suelo. Aquí los BLOBS están ordenados para ser enviados al EAV.

Tabla 9. Resultados del filtrado de BLOBS por distancia y área normalizada

Id	Área(px)	Perímetro (px)	Rectángulo delimitador (x,y,w,h)	Centro Geométrico (x,y)	Distancia (mm)
1	27532	748	(347, 307, 194, 153)	(412, 386)	1719
2	205083	2908	(1, 1, 619, 431)	(329,181)	3510
3	13538	818	(1, 397, 273, 63)	(145, 432)	2528

El tiempo de procesamiento que se emplea en la Etapa de procesamiento y análisis es de 12 fps.

6.3 Virtualización y reproducción del EAV

En la Figura 40, se observa el proceso de generación, sintetización y modulación de una señal acústica para una fuente de sonido virtual activada.

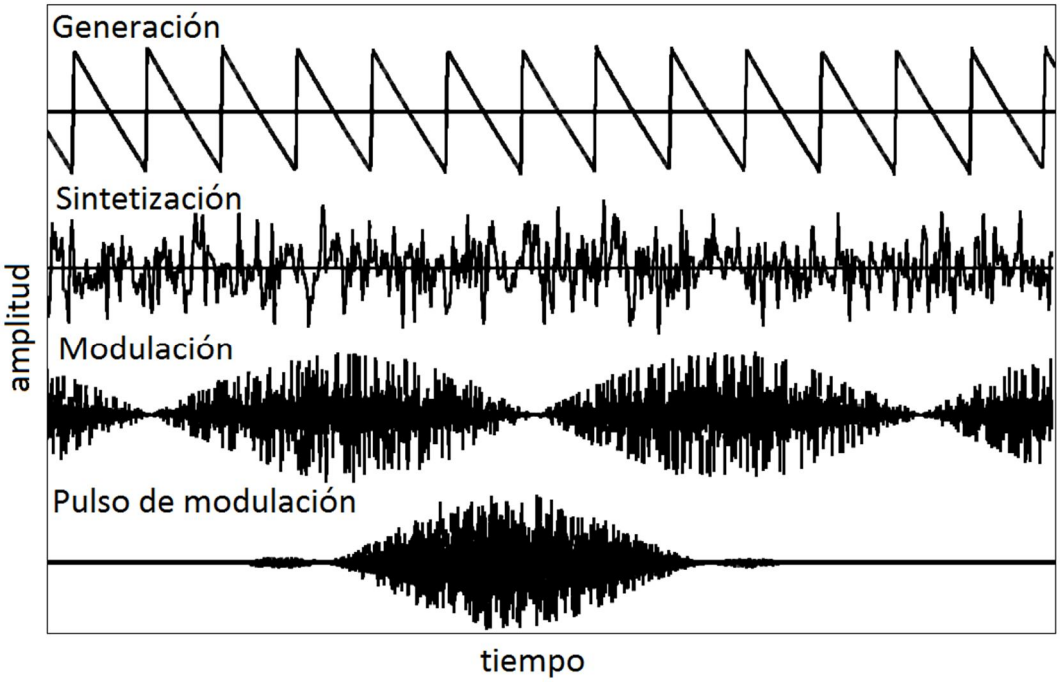

Figura 40. Resultado del sintetizador aplicando las FTA para una fuente de sonido.

Las fuentes de sonido tienen una de las frecuencias fundamentales presentes en la Tabla 10 para el proceso de sintetización de la señal. Para la frecuencia de modulación, a la fuente de sonido se asigna uno de los tres valores posibles de 3, 19, 53 Hz, según la relevancia de objeto.

Tabla 10. Escala de frecuencias utilizadas (Hz)

f1	f2	f3	f4	f5	f6	f7
220	329	493	739	1108	1661	2489

El tiempo que emplea la etapa de virtualización y reproducción por las 105 fuentes de sonido, en promedio es de 483 fps.

6.4 Pruebas del prototipo

Se realizaron tres tipos diferentes de pruebas: ubicación de objetos virtuales, ubicación de objetos reales y de movilidad. Para realizar las pruebas de desempeño del prototipo de laboratorio se estableció una secuencia de entrenamiento de aproximadamente 30 minutos. El entrenamiento consistió en familiarizar a la persona en la interpretación de la lateralidad (derecha – izquierda), verticalidad (arriba – abajo) y distancia (cerca – lejos) tanto para la posición de un objeto, como para dos y tres objetos. La lateralidad en virtud del comportamiento de la espacialización sonora, la verticalidad o altura por cambios en las frecuencias, y la distancia mediante cambios en la amplitud de los sonidos. La cantidad de pulsos en un período de reproducción permite detectar el número de objetos y para diferenciarlos entre sí, se introducen cambios en el timbre de las señales por efecto de la modulación.

6.4.1 Primera prueba

Para esta primera prueba se divide el mapa de profundidad en nueve zonas: arriba – izquierda (AI), arriba (A), arriba – derecha (AD), izquierda (I), centro (C), derecha (D), abajo – izquierda (AbI), abajo (Ab), abajo – derecha (AbD). Se ubica un objeto virtual de forma aleatoria en una de las anteriores zonas y la persona debe indicar la posición. Los resultados de esta prueba para un objeto se muestran en la Tabla 11. Los resultados para dos objetos aparecen en la Tabla 12, y para tres, en la Tabla 13. La prueba se realizó a 6 personas entre los 18 y 60 años, cinco de ellas no tienen ninguna limitación visual a las cuales se les cubrió los ojos con un antifaz, mientras que una persona presenta ceguera total debido a un desprendimiento de retina.

Tabla 11. Resultados de la primera prueba. Detección de un objeto

Respuesta	**Ubicación correcta**								
	AI	A	AD	I	C	D	AbI	Ab	AbD
AI	12	1		2					
A	2	13			4				
AD		2	16						
I	1			16			1		
C		1			14	1		1	1
D						15			
AbI							17		
Ab							1	13	1
AbD									14
TOTAL (%)	80	76	100	89	78	94	89	93	87

En promedio, las personas usando el sistema completo presentan un 87% de acierto en la detección de la ubicación de un objeto.

Tabla 12. Resultados de la primera prueba. Detección de dos objetos

Respuesta	**Ubicación correcta**								
	AI	A	AD	I	C	D	AbI	Ab	AbD
AI	15	3							
A		13	1		1				
AD			12			1			
I	1			13	1		3		
C		1			9				1
D						14			1
AbI				1			12	1	
Ab								10	
AbD						1		1	8
TOTAL (%)	94	76	92	93	75	87	80	83	80

En promedio, con este sistema se obtuvo un 85% de acierto en la detección de la ubicación de dos objetos.

Tabla 13. Resultados de la primera prueba. Detección de tres objetos

Respuesta	**Ubicación correcta**								
	AI	A	AD	I	C	D	AbI	Ab	AbD
AI	6	2							
A		5							
AD			6						
I	1			5			1		
C					4				1
D						7			1
AbI							5		
Ab					1		1	4	
AbD						1			5
TOTAL (%)	86	71	100	100	80	87.5	71	100	71

En promedio, se obtuvo un 84% de acierto en la detección de la ubicación de tres objetos.

6.4.2 Segunda prueba

La segunda prueba consiste en la detección de la cantidad, ubicación y orden (del más cercano al más lejano) en el que se encuentran los objetos reales (cajas) respecto a la persona dentro de una habitación. Los resultados se presentan en las Tabla 14, 15 y 16.

Tabla 14. Resultados de la segunda prueba. Cantidad de objetos

Respuesta	**Cantidad correcta**		
	1	2	3
1	32		
2		33	1
3			33
TOTAL (%)	100	100	97

En promedio, las personas detectaron la cantidad de objetos presentes con un acierto del 99%

Tabla 15. Resultados de la segunda prueba. Orden de los objetos

Respuesta	**Orden correcto**		
	1°	2°	3°
1°	99		
2°		66	1
3°		1	32
TOTAL (%)	100	99	97

En promedio, las personas detectaron el orden de los objetos presentes con un acierto del 99%

Tabla 16. Resultados de la segunda prueba. Ubicación de los objetos

Respuesta	**Ubicación Correcta**		
	Izquierda	Centro	Derecha
Izquierda	68	1	
Centro	3	55	1
Derecha	1		70
TOTAL (%)	94	98	99

En promedio, las personas detectaron la ubicación de los objetos presentes con un acierto de 97%.

6.4.3 Tercera prueba

Esta prueba constituye un avance en el tema de la movilidad. La prueba consiste en que la persona recorra un pasillo con los ojos vendados de manera que pueda eludir una serie de obstáculos sin colisionar. En esta prueba no se contó con

entrenamiento específico previo. En la Tabla 17 se presentan los resultados del número de colisiones totales por recorrido.

Tabla 17. Resultados de la tercera prueba. Movilidad con el prototipo

Cantidad de recorridos	Número de obstáculos	Número de colisiones	Probabilidad de eludir un obstáculo
4	2	3	63%
8	3	6	75%
12	4	19	60%

En promedio, la probabilidad de que una persona eluda un obstáculo durante un recorrido es del 65%. Esta prueba depende en gran medida de la persona que utilice el dispositivo. La persona con mejores resultados obtuvo una probabilidad de eludir obstáculos del 94%, y la persona con resultados no tan favorables obtuvo una probabilidad de eludir obstáculos del 31%.

La detección de suelos durante la tercera prueba permitía guiar a la persona y brindar mayores posibilidades para encontrar un camino libre de obstáculos. Para ello, se le asignó un sonido diferente del resto de objetos y sin pulsaciones. La distancia máxima de la región de suelo que se podía detectar varía entre los 2 metros y depende de la posición y dirección en la que se encuentre el Kinect respecto al suelo.

6.5 Análisis de los resultados

Los resultados obtenidos a la fecha permiten determinar el desempeño general y específico del sistema implementado en un prototipo de laboratorio. En términos generales, el sistema muestra un buen desempeño para que una persona interprete la ubicación de máximo tres objetos en el entorno, con un límite inferior

de 84%, y un desempeño aceptable del 65% en promedio para que una persona pueda eludir obstáculos durante un recorrido.

En cuanto al desempeño de las tres etapas: 1) adquisición, 2) procesamiento y análisis, y 3) virtualización y reproducción. En la etapa de adquisición, el tiempo de captura de las imágenes de mapas de profundidad de 30 fps, se puede considerar como un óptimo desempeño para obtener mapas de profundidad, si se compara con las otras técnicas existentes de visión estereoscópica, luz estructurada y triangulación láser, las que difícilmente proveen mapas de profundidad a esa velocidad.

En la etapa de procesamiento y análisis, el tiempo de procesamiento promedio de 12 fps es un óptimo desempeño para sistemas de visión por computador. Los resultados de la etapa de procesamiento permiten observar que el sistema detecta la presencia de mínimo tres objetos en el entorno en un rango de 0.5m y 4.0m, debido al rango de funcionamiento del Kinect. A los objetos detectados se les puede realizar la extracción de características y seguimiento, y dichas características escogidas son las adecuadas para realizar el seguimiento y extraer su posición cartesiana relativa a la persona en el entorno. Es posible añadir nuevas características con el fin de mejorar y añadir nuevas funcionalidades al EAV.

En la etapa de reproducción y virtualización, el número de fuentes de sonido virtuales es suficiente para que una persona detecte la ubicación de tres objetos presentes en el entorno con un acierto del 84 %. Esta etapa tiene alto grado de flexibilidad, para realizar nuevos estudios en el desempeño y funcionamiento, en particular para mejorar la probabilidad del 65% al eludir obstáculos durante la movilidad a través de un recorrido con obstáculos. Dentro de las posibles modificaciones a estudiar se citan los coeficientes de Fourier, las frecuencias de modulación que cambian el timbre de las señales, la escala de frecuencias ajustable a la persona que utilice el sistema según su percepción acústica. También se puede modificar la espacialización sonora, si se cambia la escala de

distancias del EAV al ampliar la lateralidad del sistema. Finalmente se puede ajustar el tiempo de pulsación de las señales para incrementar la frecuencia de reproducción que indica la posición de los objetos, y aumentar la probabilidad de detectarlos durante la movilidad para eludirlos. Adicionalmente, una de las ventajas respecto a los sistemas del Capítulo 2, es el corto tiempo de entrenamiento y aprendizaje.

El sistema completo tiene una gran cantidad de variables y parámetros susceptibles de ajustar según el propósito y uso que se le quiera dar. En ésta investigación, el principal asunto a tratar fue la detección y ubicación de objetos en el entorno que no producen sonido por sí mismos. El sistema se ajusta correctamente para ese propósito; detecta objetos por medio de imágenes de mapas de profundidad y genera un EAV que le permite a una persona ubicar hasta tres objetos presentes en la escena.

7. CONCLUSIONES

Se evidencia que es posible generar un Espacio Acústico Virtual a partir de imágenes de mapas de profundidad capturadas por un dispositivo óptico como el sensor Kinect.

Los avances en la tecnología de visión por computador en cuanto a sensores ópticos de profundidad como el Kinect; técnicas de procesamiento de imágenes como la segmentación por ubicación de mínimos, correlación de imágenes, etiquetado de regiones por contornos, entre otras; y en herramientas software como OpenNI y OpenCV, permiten obtener, procesar y analizar mapas de profundidad para detectar objetos presentes en el entorno y extraer sus ubicaciones en tiempo real (10 fps).

El sistema auditivo humano es un buen sustituto sensorial para el sentido de la vista en la detección y ubicación de objetos en el entorno cercano, siempre y cuando se escojan adecuadamente los parámetros acústicos para las FTA que generan las fuentes de sonidos virtuales del EAV. Es posible que personas entre los 18 y 60 años de edad realicen mediante el canal auditivo, la identificación y ubicación de hasta tres objetos en un determinado entorno. Igualmente, es posible detectar el suelo a través de imágenes de mapas de profundidad y transmitir la información de forma audible para que una persona la interprete y realice el proceso de movilización.

8. TRABAJOS FUTUROS

Esta investigación constituye un avance en cuanto a la detección de objetos por medio de un EAV a partir de imágenes de mapas de profundidad, que puede ser mejorado en aspectos de inteligencia artificial que permitan reconocimiento de objetos, de manera que si una persona desea saber cuál es el objeto detectado, active el modo de reconocimiento de objetos, e identifique si es una persona, pared, mesa, caja, entre otros.

Se puede continuar en el estudio para establecer las configuraciones de los parámetros acústicos y las FTA adecuadas para mejorar la detección de los objetos en el entorno, aumentar la cantidad de objetos detectables y aumentar la probabilidad de eludir obstáculos durante la movilidad. También, es necesario adelantar los estudios respecto a la evolución de una persona con tiempos prolongados de entrenamiento y exposición con el sistema.

Otro campo de investigaciones es en términos de la portabilidad hardware del sistema, en la implementación de un prototipo embebido que incluya el dispositivo óptico en un par de gafas y la unidad de procesamiento en un dispositivo más portable como un celular o una *tablet*.

Se puede complementar la etapa de procesamiento para detectar objetos particulares como escaleras, andenes, escalones, andamios, huecos, y obstáculos similares y asistir a personas en situación de discapacidad visual para que aumenten la probabilidad de eludirlos.

Además del modo de detección de objetos en el entorno, se pueden implementar nuevos modos de funcionamiento en el EAV que brinden asistencia más específica según las necesidades de la persona. Podrían añadirse los modos de movilidad, lúdica, aprendizaje, identificación de la forma de los objetos, entre otros.

BIBLIOGRAFÍA

Bach-y Rita, P., Collins, C., Saunders, F., White, B., & Scadden, L. (1969). *Vision substitution by tactile image projection.* Nature.

Bach-y Rita, P., Kaczmarek, K., Tyler, M., & García-Lara, J. (1998). *Form perception with a 49-point electrotactil stimulus array on the tongue.* Journal of Rehabilitation Research Development.

Bradski, G., & Kaehler, A. (2008). *Learning OpenCV.* United States of America: y O'Reilly Media, Inc.

Bristow-Johnson, R. (1996). *Wavetable Synthesis 101, A Fundamental Perspective.* 101st AES Convention (Los Angeles, California): Audio Engineering Society (AES).

Cheng, C., & Wakefield, G. (2001). *Introduction to Head-Related Transfer Functions - HRTF-: Representations of HRTFs in Time, Frecuency and Space.* University of Michigan.

Correa, L. M. (2012). *Documento Territorial Cauca N° 10.* INSTITUTO NACIONAL PARA CIEGOS.

Escalera, A. d. (2001). *Visión por Computador. Fundamentos y Métodos.* Madrid: Prentice Hall.

Freedman, B., Shpunt, A., Machline, M., & Arieli, Y. (2010). *Patent No. 20100118123.* PRIME SENSE LTD .

Gelfand, S. (2010). *Hearing: An Introduction to Psychological and Physiological Acoustics.* London: Informa Healthcare.

Grupo de Investigaciones y Desarrollo en percepción del espacio usando sonidos con aplicación para personas ciegas. (2002). Recuperado el 19 de Octubre de 2011, de http://www.iac.es/proyecto/eavi/

INSTITUTO NACIONAL PARA CIEGOS. (Octubre de 2006). *Estadisticas de Discapacidad Visual en Colombia.* Recuperado el Septiembre de 2012, de www.inci.gov.co/ftp/informacion_estadistica_plv_2005

Lanman, D., & Taubin, G. (2009). *Build your own 3D scanner: 3D photography for beginners.* Brown University: SIGGRAPH 2009 Course Notes.

López, H. (2011). *Detección de objetos con cámaras en movimiento.* Grupo de Neurocomputación Biológica: Universidad Autónoma de Madrid.

Meijer, P. (1992). *An Experimental System for Auditory Image Respresentations.* IEEE Transactions Biomedical Engineering.

OpenAL. (s.f.). *Creative Labs*. Recuperado el 17 de Mayo de 2012, de http://connect.creativelabs.com/openal/default.aspx

OpenCV. (s.f.). Recuperado el 10 de Febrero de 2012, de http://www.opencv.org/

OpenNI. (2010). *Open Natural Interface*. Recuperado el 25 de Enero de 2012, de http://www.openni.org/openni-sdk/

ORGANIZACIÓN MUNDIAL DE LA SALUD. (2012, Junio 5). *Ceguera y discapacidad visual*. Retrieved Septiembre 2012, from http://www.who.int/mediacentre/factsheets/fs282/es/index.html

ORGANIZACIÓN MUNDIAL DE LA SALUD. (Junio de 2012). *Prevention of blindness and visual impairment*. Recuperado el 5 de Septiembre de 2012, de <http://www.who.int/blindness/actionplan/en/index.html>

Peris-Fajarnés, G. (2009). *Congnitive Aid System for Blind People FINAL ACTIVITY REPORT.* Universidad Politécnica de Valencia.

Praderas, S., Ortigosa, N., Dunai, L., & Peris-Fajarnés, G. (2009). *COGNITIVE AID SYSTEM FOR BLIND PEOPLE (CASBLiP).* Valencia, España: Centro de Investigación en Tecnologías Gráficas y Universidad Politécnica de Valencia.

Rodriguez-Ramos, L., Chulani, H., Diaz-Saco, L., Sosa, N., González-Mora, J., & Rodriguez-Hernandez, A. (2006). *A Image and Sound proccesing for the creation of a virtual acoustic space for the blind people.* Instituto de Astrofísica de Canarias y Universidad de la Laguna.

ROS. (s.f.). Recuperado el 28 de Enero de 2012, de http://www.ros.org/wiki/kinect_calibration/technical

Sáez, D. (2010). *Orientación y Movilidad*. Recuperado el 5 de Octubre de 2012, de Sense: http://www.sordoceguera.org/vc3/para_maestros_profesionales/orientacion_movilidad.php

Wikipedia. (11 de Octubre de 2012). *Streaming*. Recuperado el 15 de Octubre de 2012, de http://es.wikipedia.org/wiki/Streaming

Wikipedia. (9 de abril de 2013). *Realidad aumentada*. Recuperado el 9 de abril de 2013, de http://es.wikipedia.org/wiki/Realidad_aumentada

Zöllner, M., Huber, S., Jetter, H., & Reiterer, H. (2011). *NAVI - A Proof-of-Concept of a Mobile Navigational Aid for Visually Impaired Based on the Microsoft Kinect.* University of Konstanz.