Utilización de tecnología de multimedia con personas discapacitadas

HiperAudio: un Sistema Hipermedial para No Videntes

Mauricio Fabián Lumbreras

Director: Lic. Gustavo H. Rossi

Trabajo de grado de Licenciatura en Informática
Departamento de Informática - Facultad de Ciencias Exactas
Universidad Nacional de La Plata
ARGENTINA

Diciembre de 1995

Indice

Capítulo 2 : Acústica y sicoacústica del espacio auditorial virtual

Capítulo 3 : Procesamiento digital de señales y sistemas

Capítulo 4 : Hipermedia

Capítulo 5 : Introducción a los Displays Acústicos

Capítulo 6 : Introducción a hipermedia acústica tridimensional

Capítulo 7 : HiperAudio

Agradecimientos

Cuando a final de 1992 Gustavo Rossi me propuso trabajar en un proyecto que involucrara el uso de tecnología multimedial en personas discapacitadas, no tenía la más remota idea que todo ello llevaría a la realización de este trabajo. Esta idea que propuse en su momento, fue lentamente tomando forma a medida que recababa información. La idea fue madurando y así fue publicada en las conferencias de ACM Hypertext 93 y CHI 95 entre los lugares de más renombre. Estos hechos fueron motivos para pensar que esa idea exótica no sólo me parecía bien a mi, sino a personas que están en otra esfera académicas. Pero para poder concretarla definitivamente quiero expresar mi agradecimiento a:

Mariano Barcia, por su inestimable tiempo y ayuda en la implementación del editor de HiperAudio, sin el cual el trabajo hubiera sido muy difícil.

Gustavo Patow, con el cual discutimos tópicos acerca de sonido 3D.

Fernando Das Neves, por sus charlas críticas que me sirvieron para plantear nuevas ideas.

Barry Arons, que me dió su tesis de doctorado la cual me sirvió de base para plantear mis ideas y como modelo para imitar por su claridad y organización de la información.

Alejandra Garrido, que me facilitó el formato para poder presentar prolijamente este trabajo

Ramiro Gonzalez Maciel, el cual me ayudó con sus voz a generar el prototipo del HiperaAudio.

Mónica, mi profesora de inglés que me bancó en la correción de mis papers.

Polo y mi mamá, los cuales incondicionalmente me apoyaron a full en todo lo que me implicó este trabajo.

Mi papá, que hoy ya no está pero que le hubiese gustado ver esto terminado.

Pero fundamentalmente mi agradecimiento va a Gustavo Rossi, quién siempre me apoyó para poder viajar a contar mis ideas, se entusiasmó y gestionó los fondos para comprar el hardware necesario sin saber si el objetivo final sería alcanzado, además de nuestras "micro-discusiones" en las cuales discutiamos las ideas de este trabajo.

El trabajo fue largo, pero puedo asegurar que la experiencia adquirida en mis viajes producto de esta idea, no tienen precio. La dedicación a este documento me obligó a aprender una manera de plasmar las ideas en el papel, de organizar y de llevar a cabo una idea compleja. Pienso que la idea de trabajo de grado está cumplida.

La próxima fase, ya no depende tanto de mi. La potencial distribución de un sistema como el propuesto a todos los que le haga falta será una tarea compartida. Es así que luego de casi 3 años de trabajo puedo decir que la primera fase del trabajo está terminada.

Resúmen

Capítulo

0

El acceso a la información es uno de los aspectos más importantes hoy en día para el desenvolvimiento de un individuo en la sociedad. El advenimiento de la tecnología digital introduce una forma de soportar la información, sin papel ni tinta. Este medio electrónico hace de soporte no sólo de texto crudo, sino que se integra éste con gráficos, video y sonidos. Además, esta integración involucra un método de estructuración diferente al que existe en los medios impresos: la organización es hipermedial, introduciendo una manera diferente de navegar y explorar la información. También es ampliamente conocido que las aplicaciones hipermediales, en particular aquellas presentadas en CD-ROM, serán preponderantes en ciertos dominios, tal como los educativos.

Desafortunadamente, las metáforas de presentación y acceso hacen una profusa utilización de gráficos, imágenes dinámicas, íconos, etc., sin tener en cuenta el gap que producen estas entre usuarios normales y ciegos.

Nuestra aproximación plantea una manera especial de representar la información respetando el modelo hipermedial. Por ello se presenta en este trabajo una metáfora conversacional que permite el acceso a una base de información hipermedial, presentada esta enteramente a traves de sonido tridimensional. El uso de esta tecnología intenta explotar el sentido del oído al máximo, pues el destinatario final del sistema será un usuario ciego.

Usando el paradigma hipermedial como modelo subyacente y tecnología de sonido 3D como medio expresivo de información, mostraremos como se puede construir un sistema, de tal manera que este pueda ser utilizado por una persona ciega. La idea básica plantea al usuario como un moderador de una conversación entre múltiples locutores. Bajo el contexto de la metáfora conversacional propuesta, discutiremos como el usuario obtiene información, la maneja y controla su flujo. Por medio de una variante de la metáfora de rooms, analizaremos como representar acústicamente la arquitectura estática del ambiente virtual en el cual el usuario navega.

Estas decisiones imponen nuevos tópicos no explorados en la bibliografía, tal como modos de navegación espacial en ambientes sin pistas visuales y guidelines para el diseño de displays virtuales acústicos. Además, la interacción con el sistema se lleva a cabo interactuando con un guante para aplicaciones de realidad virtual, lo cual impone nuevos desafíos en la manipulación de entidades virtuales acústicas sin pistas visuales.

Estas ideas previas son las que imponen y generan el trabajo original que se expresa en el sistema desarrollado, el cual se denomina HiperAudio.

A grandes trazos se explicarán los items a tener en cuenta en el diseño de un sistema de HiperAudio; el cual involucra una revisión de las tecnologías de información para no videntes, acústica y sicoacústica de la percepción de sonidos espaciales, procesamiento de señales y lineas de investigación en tecnología hipermedial entre otros tópicos.

Finalmente se mostrarán la herramientas para la edición del HiperAudio, características de la implementación y experiencias de cada una de las fases del proyecto.

Medios de Soporte de Información para No Videntes

Capítulo

1

Introducción

Cómo un no vidente adquiere conocimiento? La idea de este capítulo es revisar de que manera se puede expresar información sin utilizar la vista y cómo un sistema de cómputo básicamente puede ser accedido por un ciego. Cada item servirá para imaginar que características ofrece cada uno de ellos como soporte o medio de expresión de información.

Revisaremos brevemente los trabajos más relevantes que hacen énfasis en sistemas o modelos dónde la interacción y presentación de información carece de representación visual. Con lo anterior en mente, tendremos un contexto adecuado para situar el diseño finalmente propuesto.

El problema: escenarios actuales para el soporte de información

La falta del sentido de la vista obliga con referencia al acceso a información, a sustituirlo a este por el tacto y el oído. Mientras que la suplencia por medio del tacto la hace la misma persona directamente, la sustitución por medio del oído exigía hasta ahora la colaboración de otra persona que actuara como lector, con lo que se limitaba la autonomía del ciego. Actualmente la técnica, mediante el libro hablado y el sintetizador de voz, ha resuelto parcialmente este inconveniente y pone al oído en el mismo plano que el tacto, con respecto a la capacidad de facilitar la autonomía personal.

Antes de diseñar nuestro sistema deberiamos examinar brevemente que medios de información existen hoy para no videntes:

• El texto braille es aceptado pero posee algunas desventajas, tal como la dificultad en la búsqueda de información, el tamaño de los libros (debido a la baja densidad inherente de los caracteres braille), el problema de mantener la información actualizada (debido al coste de reimpresión), el problema que presenta aprender esta nueva forma de describir el alfabeto (solo el 10% de ciegos maneja braille [Vanderheiden 92]).

• Los cassettes de audio son otro medio. Aunque su costo es bajo y permite secuenciamiento, no soportan interacción del usuario, no existen presentaciones especiales de su contenido para buscar informacion en ellos o algún tipo de computación incluida.

Veremos a continuación como funcionan cada una de las tecnologías que existen para el soporte de información para no videntes. Además revisaremos como un usuario ciego puede acceder a un sistema de cómputo, el cual potencialmente es una fuente de información.

Información escrita

La lectura directa por el tacto se ha ensayado desde hace mucho tiempo. El español Francisco Lucas (1517) publicó su invención de letras móviles de madera en relieve para que pudieran ser leídas por ciegos. Valentín Haüy (1745-1822) educador de ciegos, observó que uno de sus aprendices podía distinguir por el tacto algunas letras accidentalmente grabadas en una cartulina, por la presión de los tipos en la prensa. Haüy imprimió posteriormente libros en relieve usando tipos modificados, pero existía la dificultad en discriminar los caracteres y además un usuario no podía escribir con este método, pues era difícil con un punzón describir la forma de una letra sin una guía.

Louis Braille (1809-1852), ciego desde los tres años, tuvo conocimiento del método desarrollado por un capitán de apellido Barbier, el cual enviaba mensajes realizados en relieve para poder ser leídos en la noche sin ayuda de la luz. El método codificadaba las letras en 12 puntos y rayas sobre un papel, los cuales eran decodificados por el tacto. Braille redujo este código a 6 puntos dispuestos en dos columnas de tres.

El aumento de la eficacia en lectura y escritura que se obtiene por este método, es debido a que las letras, constituidas por trazos diversos, han sido sustituídas por conjuntos ordenados de un elemento único a reconocer: el punto.

En la figura 1.1 se ve que el alfabeto braille consta de "células" conteniendo seis puntos con posiciones numeradas del 1 al 6. De allí surge que existen 64 posibles combinaciones de caracteres braille, algunas de las cuales se utilizan para letras, otras para signos de puntuación y otras como prefijos para indicar mayúsculas o dígitos numéricos.

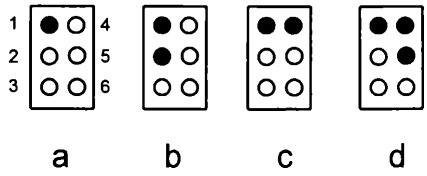


Fig 1.1: Representación de algunos caracteres alfabéticos en la representación diseñada por Braille. El punto negro indica un relieve saliente en el papel.

Las hojas de papel o plástico grabadas con estos caracteres constituyen para las personas ciegas elementos permanentes de lectura, tales como los libros tradicionales funcionan para los videntes.

El terminal braille

De la misma manera que existen módulos alfanuméricos de matrices de puntos (tal como aparecen en las calculadoras) capaces de representar a la vista caracteres con los que se puede formar un mensaje transitorio, también se han desarrollado para la lectura braille módulos mecánicos con una matriz de 6 u 8 puntos en relieve. Estos puntos son retraibles por la acción de pequeños solenoides o elementos piezoeléctricos.

Estos módulos mecánicos se conocen como *células braille* y un conjunto de estos dispuestos en línea constituye una línea braille que puede tener 20, 40 u 80 elementos. Esta línea usada como terminal de la computadora, es capaz de reproducir en braille mediante software y la interfaz adecuada, una línea de texto convencional. El usuario lee esa línea pasando el dedo sobre ella como si se tratara de una línea impresa. Una vez leída, un nuevo conjunto de caracteres ocupa el lugar de los anteriores, y de esta manera se prosigue hasta completar un texto dado.

Acceso a la información por medio de voz sintetizada

La salida táctil no es la única para transmitir a la persona ciega la información de textos codificados en memoria de la computadora. Los sintetizadores de voz permiten a la información escrita hacerla inteligible a través del oído.

El hardware suele constar de tarjetas que se adicionan a la computadora, diseñadas usando chips dedicados o por cierto software que utiliza alguna placa de sonido estandard. En ambos casos estos sistemas son capaces de generar un conjunto de fonemas, que permite la pronunciación de palabras mediante el uso de reglas gramaticales y con ayuda de diccionarios. En los sistemas de primera

genración es común obtener una voz poco expresiva y con un tono "robótico". Técnicas mas moderanas permiten simular las condiciones del tracto vocal y por lo tanto son capaces de dar entonación y acento, consiguiendo así casi una calidad humana.

Los sintetizadores de voz tienen la capacidad de variar la velocidad de pronunciación desde una velocidad lenta para pasajes difíciles hasta otra rápida equivalente a la que pueda tener una persona vidente leyendo. También son capaces de deletrear palabras en caso de revisión ortográfica y acrónimos. Pueden también tener timbre masculino o femenino.

El sintetizador de voz puede leer una pantalla en modo texto, pero no es suficiente para manejar programas y formatos complicados de pantalla. Para conseguirlo es necesario utilizar además un software lector de pantalla (*screen reader*). Este software permite recorrer la pantalla y asi acceder a cualquier software basado en modo texto. De esta manera, un no-vidente puede aprovechar gran parte de las posibilidades de una computadora.

Acceso genérico a GUIs

El advenimiento y proliferación de GUIs durante la última década y la reciente introducción de conceptos multimediales en la interacción hombre-máquina trae consigo nuevas posibilidades, pero también muchos problemas para usuarios ciegos con referencia al acceso a sistemas de cómputo e información digitalmente soportada. Resultados de encuestas indican que la mayoría de las compañias productoras de software (93%) producen software para GUIs y sólo un porcentaje de ellas (27%) planea vender versiones basadas en interfases de texto [Blenkhorn 92]. Considerando MS Windows hasta la version 3.1, mas de diez millones de copias han sido vendidas de acuerdo a Microsoft.

Hasta hace poco, las GUIs fueron totalemente inaccesibles para usuarios ciegos. Estas interfases han sido diseñadas para explotar las capacidades visuales de videntes por medio de complicadas representaciones pictóricas, haciendo para los *screen readers* la tarea sumamente difícil. Actualmente existen adaptaciones para Macintosh (Outspoken) y OS/2 (Screen Reader/PM) que basadas en síntesis de voz permiten describir la pantalla. System 3 es otra adaptación para Macintosh, el cual integra un *pointing device* , denominado Optacon y un sintetizador de voz.

Estos intentos iniciales no dejan claro que elementos de una GUI pueden ser exitosamente adaptados a una nueva modalidad. Mientras el texto puede ser

presentado con voz o braille, la interacción con objetos gráficos complejos como menús pop-up, cajas de diálogo y scroll bars pueden ser mejor descriptos a través de una integración consistente entre audio y tacto. Sin embargo, los screen readers que aparecen actualmente tienen una capacidad limitada para soportar la interacción con estos objetos gráficos, a pesar de usar combinaciones de braille, voz y sonido. Además, la interacción por medio de un pointing device como el mouse no es usualmente soportada.

Dado el estado del arte de los lectores de pantalla, dos proyectos intentan generar un acceso genérico a aplicaciones basadas en GUI. Ellos son el proyecto GUIB [Weber 93], auspiciado por el cosorcio TIDE de la CEE; y el otro el proyecto Mercator, desarrollado en el Instituto de Tecnológico de Georgia, USA [Mynatt 92]. Brevemente revisaremos como funcionan cada uno de ellos.

Proyecto GUIB

El proyecto GUIB se basa fundamentalmente en el uso de un dispositivo especial que integra líneas braille, sonido stereo, un panel sensitivo a la presión y teclas para manipulación del cursor [Weber 93]. La idea es remapear los eventos y simbología de cada artefacto gráfico de Windows a este dispositivo. Por medio de las lineas braile, el usuario es capaz de explorar la pantalla . Teclas especiales permiten activar un sintetizador de voz que describe el estado de la interfaz y lee caracteres ingresados. Presionando y moviendo la yema del dedo sobre un panel o *pad* permite direccionar el mouse de manera espacial en forma absoluta. Por medio de los auriculares es posible representar cierta información espacial, manipulando la intensidad del sonido en cada auricular. Con respecto al software, una aplicación especial, captura todos los eventos que genera una aplicación Windows, y los remapea al nuevo dispositivo; por ej: generando una descripción textual de cada ícono y remapeando ciertos eventos a sonidos, representando el movimiento del mouse como sonido de pasos sobre diferentes superficies dependiendo la posición relativa del cursor y la ventana que está debajo. Entre otras cosas, la forma del cursor es reproducida con un sonido de goma elástica si el cursor está en el borde de la pantalla, o con un tic-tac de reloj cuando el cursor está con forma de reloj de arena.

Actualmente se están realizando test experimentales y no está comercializando hasta la fecha.

Proyecto Mercator

La idea del proyecto Mercator [Mynatt 92] es similar a la de GUIB, pero sin utilizar dispositivos especiales. La plataforma de implementación de esta

aproximación es cualquiera que corra el protocolo X (X-Windows). La información referente a la interfaz gráfica es modelada como una estructura de árbol la cual representa los objetos gráficos en la interfase (botones, menús, áreas de texto, etc.) y la relación jerárquica entre estos objetos. De ésta manera un ciego interactúa con el sistema independientemente de su representación gráfica. El contenido de la interfaz es expresado a través de voz sintetizada y *nonspeech audio* (audio que no representa voz). A través de diferentes características del sonido, el sistema expresa atributos de la interfaz: por ej.: un menú desabilitado se expresa con el sonido original del menú, pero pasado por un filtro pasabajos, reproduciendose el sonido sin brillo, dando idea que está desabilitada esa opción.

En un nivel simple, el usuario navega en la interfaz Mercator, cambiando su posición en el árbol representado, usando el teclado. Cada movimiento expresado con el teclado, permite moverse en el árbol, y a través de pistas de audio, el sistema expresa la naturaleza de cada nodo del árbol.

Si bien lo interesante de estos enfoques es el uso genérico de una computadora por parte de un no vidente, el enfoque planteado en HiperAudio, es más específico y se centra en una modalidad para representar información de tal manera que pueda ser accedida sin pistas visuales.

El campo de aplicación de sistemas de información enteramente acústicos esta siempre obviamente caracterizado por la falta de pistas visuales para seleccionar y navegar la información. Este detalle hace interesante el estudio de los trabajos relacionados con esta modalidad, para aplicar sus resultados asistemas de información para ciegos.

Acceso a información sin pistas visuales: Trabajos previos

A continuación y a modo de breve revisión, se describen los trabajos más relevantes relacionados con el browsing y almacenamiento de información para ser recuperada sin pistas visuales.

Hyper Phone

HyperPhone [Muller 92] es un prototipo experimental para comunicaciones de voz soportadas por computadora desarrollado en Bellcore, en el cual se provee un ambiente para interactuar con entidades denominadas *voice documents* (documentos acústicos) y aplicaciones operadas por la voz. Los tópicos explorados por el proyecto HyperPhone son los siguientes:

• Una interfaz de reconocimiento de voz para acceder a un sistema hipermedial

• Un browser para *voice documents*

• Una eventual interfaz conversacional para interactuar con ciertas operaciones bajo el contexto de un sistema de telecomunicaciones

• Una potencial aplicación para usuarios que no pueden acceder a displays o teclados

Este trabajo básicamente expresa importantes ideas acerca del modelo conversacional como metáfora de interacción usuario-computadora. Además plantea una arquitectura basada en capas para el diseño de aplicaciones hipermediales de documentos de voz, tal como capa de presentación, y capa de procesamiento y de datos , separando claramente contenido de presentación.

Además expresa los resultados que impactan al usuario tal como granularidad de la información, complejidad y ejecución de links. Se discute además la posibilidad de adicionar inteligencia a la interacción a través de agentes.

Speech Skimmer

Capturar un determinado documento grabado, tal como el que puede provenir de una conferencia o monólogo de un locutor, no sólo implica el almacenamiento y posterior reproducción. La búsqueda de información es un aspecto sumamente complejo, pues se debe optimizar el recurso más costoso, que es el tiempo. Este trabajo [Arons 93] no trata con la estructuración de información de manera hipermedial, pero exhibe técnicas para segmentar automáticamente grabaciones monolíticas en pequeños *chunks* o porciones atómicas de información. Además la eliminación de pausas y compresión temporal del sonido son incluídos como técnicas para minimizar el tiempo de escucha. De esta manera se puede obtener información con diferente grado de detalle. El autor crea una interfaz tipo panel, especialmente realizada para interactuar con documentos organizados con la técnica descripta y reporta el comentario de los usuarios. De este trabajo se desprenden técnicas para minimizar el tiempo de reproducción de un documento, sin perder inteligibilidad.

Voice Notes

VoiceNotes es una aplicación para controlar un pequeño dispositivo que se puede llevar en forma autónoma y portátil, que permite capturar, manejar y obtener lo que denominan *voice notes* o clips de audio [Stiefelman 93]. Estos clips representan ideas, cosas para hacer o pensamientos que son expresados por el usuario, y capturados espontáneamente por el dispositivo. Además de describir los problemas que ocurren en la interacción con reconocimeinto de voz, el uso de voz como medio de expresión de información, etc. también da idea de cómo debe organizarse la información para minimizar desorientación. Plantea un simple modelo jerárquico basado en categorías y notas dependientes, como el más favorable para este dominio. Interfaces modales vs. no modales, señales no vocales como feedback de interacción, conservación fundamental del recurso tiempo y estructuración de la información en forma dinámica son los puntos interesantes a tener en cuenta como legado para el sistema que se propone en esta tesis

The SpeechActs System

El uso de aplicaciones o sistemas móviles implica interacción con interfaces simples, ya que estas deben ser portátiles, fáciles de operar en cualquier contexto y que minimizan las interacciones innecesarias. Un área de gran interés es la utilización de operaciones remotas a través de líneas telefónicas, utilizando la voz como medio de expresión de comandos. De este trabajo [Yankelovich 95] se rescata la importancia de un modelo conversacional como modo de interacción con el sistema y el error que ocurre al tratar de transaladar las modalidades gráficas de interacción a una modalidad enteramente auditiva. Si bien en HiperAudio no se utiliza *a priori* reconocimiento de voz tal como lo hace este trabajo, se desprenden de él interesantes ideas asociadas a un modelo conversacional como metáfora de interacción, y cómo debe ser el manejo de errores dentro de este contexto. Estos errores surgen por malas interpretaciones del sistema a acciones iniciadas por el usuario por medio de la voz. En nuestro caso las acciones serán comandadas por un cierto dispositivo físico, pero el resultado igual es válido pues se plantean las ideas bajo el contexto de metáfora conversacional.

Hyperspeech

Tal vez este sea uno de los trabajos más interesantes relacionados con el desarrollo el sistema propuesto en esta tesis [Arons 91]. La pregunta que intenta

resolver [Arons 91] es: Cómo puede uno navegar en una bases de datos de información acústica? Si bien en HiperAudio no se utiliza reconocimiento de voz tal como sucede en Hyperspeech, es intersante observar la técnica descripta: a partir de varias entrevistas siguiendo un cuestionario previemente confeccionado, el autor digitaliza las respuestas y las asocia entre ellas, tipando los links. El usuario posteriormente a través de comandos operados con la voz, puede navegar la información. Cada comando se mapea básicamente a algun tipo de relación estática o dinámica que pueden ofrecer los nodos entre sí. Por ejemplo el usuario puede pronunciar *return* para volver atrás, o *continue* para obtener más información del locutor recientemente escuchado, o *opposing* para obtener un locutor que dice algo que se opone a lo escuchado recientemente. Los resultados más relevantes indican la dificultad en la autoría de documentos de esta naturaleza, la posibilidad de efectos de audio para indicar algo acerca de la información que se va recibir, la posibilidad de crear conversaciones virtuales entre locutores que nunca se reunieron en la realidad, la fuerte tendencia a la desorientación en un ambiente sin pistas visuales, etc. Al no utilizar metáfora espacial, deja abierto el interrogante de cómo realizar algo tal como "point & click".

Con HiperAudio se intenta solucionar este problema, y ver que ventajas y dificultades trae asociado el uso de una metáfora espacial representada con medios que inherentemente son espaciales, tal como el sonido 3D.

Por medio de una nueva tecnología que proviene de las aplicaciones de realidad virtual, podemos generar una clase especial de sonido, un sonido espacializado, genéricamente denominado sonido 3D, que permite construir cierto display virtual acústico.

Resúmen

La tecnología digital permite un amplio abanico de aplicaciones que hasta hace unos años era imposible. Tal vez el área que más impacto sufrió es la de soporte y acceso a la información. Estos sistemas que proveen acceso a vastos espacios de información, generalmente corren en una computadora sobre cierta GUI, la cual hace un profuso uso de medios de naturaleza gráfica. Además, el modelo hipermedial provee un excelente modo de acceso a la información, ya sea de manera exploratoria o instrumental. Desafortunadamente, los no videntes no pueden sacar provecho de esto. La idea fundamental es la de construir un

sistema que supla las deficiencias planteadas, sin olvidar los trabajos previos al respecto.

Hay que tener en cuenta que el diseño de interfaces y sistemas para discapacitados no sólo permitirá la integración de las personas disminuidas, sino que inexorablemente mejorará las interfaces para todos los usuarios. Los resultados que surgen del trabajo a desarrollar aquí será de importancia para cualquier sistema que se maneja con pocas o sin pistas visuales.

Acústica y Sicoacústica del Espacio Auditivo Virtual

Capítulo

2

Introducción

La representación de un ambiente auditivo que reviva los atributos espaciales de una fuente sonora es el tema en cuestión en este capítulo. Debido a que el usuario destino del sistema no contará con pistas visuales, es deseable utilizar al máximo los sentidos disponibles, siendo el del oído el que mayor ancho de banda y poder de expresividad ofrece.

Se definirá nociones básicas de las variables que intervienen en la localización espacial de fuentes sonoras en el espacio, cómo influye cada factor desde la generación del sonido hasta la percepción, que modelos describen el modo de localización de una fuente de sonido para un ser humano y como se puede simular los atributos espaciales de una fuente sonora por medio de un proceso computacional que involucra manipulación y procesamiento de señales de acuerdo a ciertos patrones.

Por qué simular un espacio auditivo?

Qué es mejor: contar lo que hay en una foto o mostrarla? Esta pregunta nos da una idea del por que es mejor recrear un ambiente acústico espacial que describirlo. Un diseñador de un sistema de sonido 3D debe tener en mente como meta ideal, la creación de un ambiente auditorial que permita al usuario final un completo control sobre la entidades del ambiente, las cuales pueden ser presentadas en forma acústica, en cualquier posición del espacio circunadante. En otras palabras, el diseñador debe poder crear un framework de entidades

acústicas de tal manera que el usuario (que puede ser el mismo diseñador) recree en su mente un contexto imaginario. Esta idea debería ser respetada en su punto extremo , en el caso de un usuario no vidente, el cual sólo puede recibir información del sistema via señales acústicas sin realimentación visual. Hay que tener en cuenta que el control de la recreación de entidades acústicas en la mente, no sólo involucra consideraciones ingenieriles de implementación de hardware basado en parámetros físicos, sino que debe tenerse en cuenta consideraciones sicoacústicas también.

Un sistema de sonido 3D es uno tal, que reemplaza o complementa los atributos espaciales de una cierta fuente de sonido. La imaginación del diseñador y las herramientas disponibles son los únicos límites a las aplicaciones que involucren manipulación espacial de sonidos.

Diferentes simulaciones de ambientes virtuales acústicos en donde involucra manipulación de sonido 3D, pueden ser recreadas a un oyente. Algunas de ellas no tienen contraparte en la realidad:

• Simulación de una condición auditorial determinada. Por ejemplo dada una grabación de un trompetista en una sala sin reberveración, se puede recrear la sensación del músico caminando por el escenario de algún auditorium previamente analizado. Este procesamiento se podría lograr para cualquier asiento del auditorium.

• Creación de un ambiente auditorial nuevo y desconocido. El trompetista anterior podría ser comprimido al tamaño de una mosca, y así volar alrededor de la cabeza del oyente.

• Transmutación entre una experiencia espacial y otra. El trompetista esta tocando a 400 metros de ud. Lentamente se va acercando a ud., haciendose mas presente hasta llegar al climax del solo terminando frente a ud. con la orquesta que se agrega a sus costados lentamente.

• Representación de una mundo auditorial determinado. A través de auriculares, se puede experimentar el solo de trompeta desde los oídos del trompetista. El usuario de un sistema de sonido 3D podría escuchar la trompeta desde el hall del auditorium, desde una silla dentro de él o detrás del escenario. El diseñador del sistema de sonido 3D tiene en este caso cada ambiente acústicamente representado previamente, de tal manera de producir la simulación adecuada.

Variables que describen atributos espaciales de una fuente de sonido

Para describir los atributos espaciales de una fuente de sonido, generalmente se utiliza un esquema egocéntrico de representación, es decir la referencia en el sistema es la cabeza del oyente o *posición del oyente.* La *fuente de sonido virtual* es el sonido que será localizado por el oyente, y la distancia lineal percibida por el oyente se denomina *distancia percibida.*

En la figura 2.1 se observan dos atributos para describir la percepción angular de la fuente de sonido virtual : *azimuth y elevación.* La perecepción del azimuth aparece particularmente robusta, debido a que los oídos de los humanos están localizados en posiciones opuestas en la cabeza, favoreciendo la detección del ángulo relativo de la fuente de sonido relativo al plano paralelo a la superficie del suelo

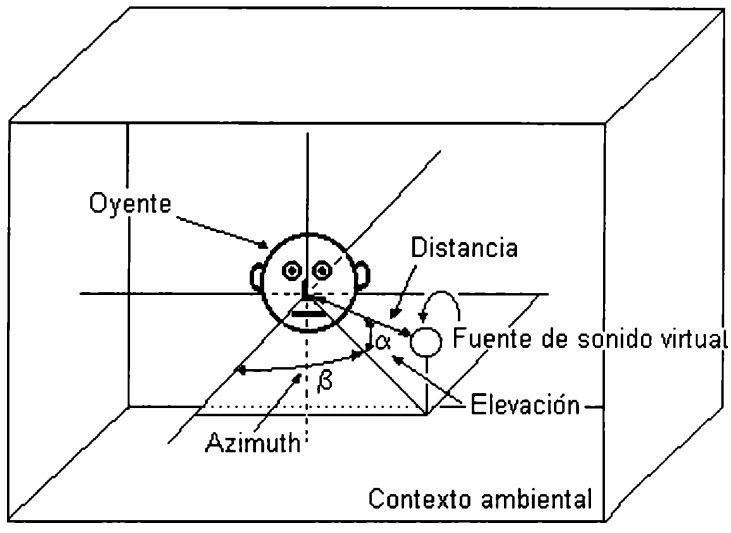


Fig. 2.1: Representación de las variables en juego en la localización de una fuente de sonido.

Normalmente el azimuth es descripto como un ángulo en grados, donde azimuth=0 y elevación=0 es la posición frente al oyente. La descripción del azimuth se puede especificar con un ángulo de 0° a 360° que se incrementa en sentido contrario a las agujas del reloj (mirando por sobre la cabeza del oyente) o como un *ángulo derecho*, que va desde el frente hacia la derecha del oyente, de 0° a 180° , y otro *ángulo izquierdo* que va desde el frente hacia la izquierda del oyente de 0° a 180°. La elevación se incrementa desde 0° (en en plano horizontal de la cabeza del oyente) hasta 90° (sobre la cabeza) y hacia abajo de la cabeza hasta -90°. Este sistema polar es cómodo para representar la posición de las fuentes virtuales de sonido, pero implementaciones reales de software utilizan la descripción de posición por medio de angulos de Euler, en términos de rotación de la cabeza, elevación de la nariz e inclinación de la cabeza (*pitch, roll & yaw*).

El azimuth y la elevación da la posición de la fuente de sonido en términos de una esfera imaginaria que rodea la cabeza del oyente, pero una descripción mas completa surge si se incluye la *distacia* de la fuente como otro atributo dimensional mas. Sorprendentemente, nuestro sentido de distancia de un sonido esta continuamente activo, pero frecuentemente imprecisamente cuantificada esta métrica. Además uno puede hablar de distancia relativa o absoluta: por

ejemplo:"...la trompeta está al doble de distancia que el piano...", o en forma absoluta:"... el mosquito está a 1 cm. del oído ...".

Otro punto a tener en cuenta es la *extensión* de la fuente de sonido: una mosca se presenta como una fuente puntual, pero una maquinaria en marcha se presenta como una fuente que genera ruido y que abarca una gran área de extensión acústica , a lo largo del azimuth y la elevación.

Finalmente existe un *contexto ambiental*. El efecto principal del ambiente en el cual se desarrolla el sonido es la *reverberación*, causada por las reflexiones repetidas de la fuente de sonido en las superficies del ambiente que encierra a ella y al oyente. De la misma manera que la luz se refleja en los objetos, el sonido hace lo propio. Este efecto es muy importante en la simulación de ambientes siguiendo este paralelo: si uno observa el rendering de una imágen sintética sin cálculos de reflejos o radiosidad, la imágen parece artificial. Si el sonido es reproducido sin reverberación, este parece como si fuera escuchado en una cámara anecoica, es decir sin ecos, no dando sensación del espacio circundante. Así el contexto circundante se presenta al oyente como un efecto dado a la fuente de sonido que genera una presentación del espacio físico ocupado por ella.

Con estos parámetros en mente, el objetivo de un sistema de sonido 3D es la recreación lo mas fiel posible de ellos para reproducir adecuadamente la experiencia auditiva a transmitir.

Modelo fuente-medio-receptor: Percepción natural vs. percepción virtual

Debe ser hecha una distinción especial entre las experiencias auditivas cotidianas y aquellas escuchadas a través de auriculares usando un sistema de reproducción de audio. La *percepción espacial natural* se refiere a cómo nosotros escuchamos sonidos espacializados normalmente, con los oídos descubiertos, con nuestra cabeza moviendose y con interacción de los otros sentidos. En la literatura aparece frecuentemente el término "hearing" (del inglés, sentido del oído o acción de escuchar). Aquí se utiliza como sinónimo la palabra *percepción*, siendo ella mas adecuada que la traducción textual.

Hay que tener en cuenta que las imágenes auditivas no estan confinadas a percepción "con dos oídos" o *binaural:* por ejemplo en una conversación telefónica uno puede experimentar pistas espaciales de la fuente emisora (por ejemplo la voz de una tercer persona que se escucha en un segundo plano). Un caso especial de percepción binaural es la *percepción espacial virtual* : esta se

refiere a la formación de imágenes espaciales acústicas sintéticas usando un sistema de sonido 3D y auriculares estereo.

El camino de transmisión desde la fuente al oyente puede ser descripta de acuerdo a un simple modelo que proviene de la teoría de la comunicación. Este modelo en su versión más simple involucra una fuente, un medio y un receptor.

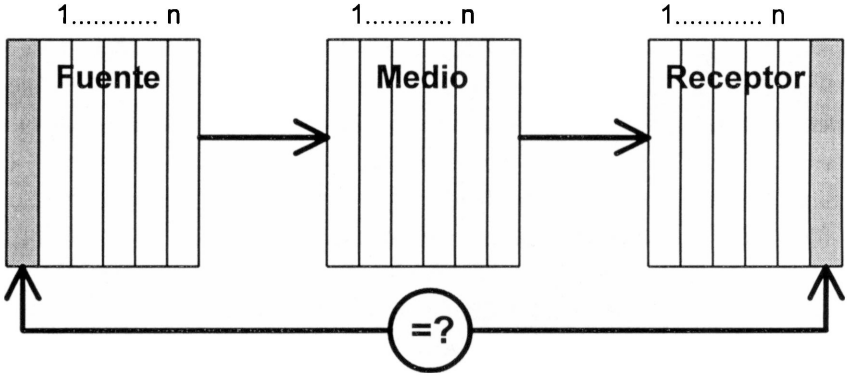


Fig. 2.2: Este es el modelo fuente-medio-receptor. Cada elemento contiene un número de transformaciones físicas, neurológicas y perceptuales (sección punteada y numerada de 1 a n) en la comunicación de la posición espacial de una fuente de sonido. Para un sistema de sonido 3D, la pregunta es: con que grado de precisión el sistema toma a la fuente e iguala al n-ésimo elemento del humano receptor, que representa el último eslabón de la cadena de percepción de la imagen virtual acústica.

Componentes involucrados en la localización

Desde el punto de vista de la sicoacústica, este modelo es interesante para ilustrar como la manipulación de una variable independiente en el sonido cambiará la imagen espacial de este. Para un sistema de sonido 3D, el propósito es dilucidar cómo una imágen auditorial deseada es precibida finalmente por el oyente.

Mientras la *fuente* puede involucrar múltiples fuentes vibratorias, es mas conveniente describir el sistema en términos de una sola fuente aislada. La percepción natural raramente incluye una fuente sola, mientras que la percepción virtual con un sistema de sonido 3D involucra el posicionamiento de un conjunto de fuentes individuales. El *medio* involucra el camino por el cual la fuente llega al oyente. En la percepción natural, este incluye el contexto ambiental (reverberación y los efectos que ofrecen los objetos físicos en la propagación del sonido); en un sistema de sonido 3D, las no-linealidades, procesamiento de la señal y los auriculares se tornan los componentes principales. Finalmente, el receptor involucra a la sicología del oyente: el sistema auditivo, desde el oído hasta las últimas fases de percepción del cerebro.

Estudiemos como funciona la percepción espacial de un sonido. La figura 2.3 muestra esquemáticamente las transformaciones que sufre el sonido desde que sale de la fuente hasta que llega al oyente a través del medio

La fuente de sonido consiste de ondulaciones en un medio elástico tal como el aire, el cual vibra de acuerdo a las vibraciones de un cierto objeto físico, tal como las cuerdas vocales dentro de la laringe o el movimiento del cono de un altoparlante. Si la fuente de sonido propaga el frente de ondas en todas las direcciones diremos que es *omnidireccional*, y por lo tanto crea un campo esférico de emisión. Si el sonido es emitido dentro de un contexto ambiental que no ofrece reflexiones - por ej. una *cámara anecoica* - el frente de ondas a partir de una cierta distancia puede ser considerado plano, significando que la presión del sonido será constante en cualquier plano perpendicular a la dirección de propagación. Pero en un ambiente no-anecoico, el sonido arriba al oyente a través de caminos directos e indirectos, tal como se ve en la figura 2.4.

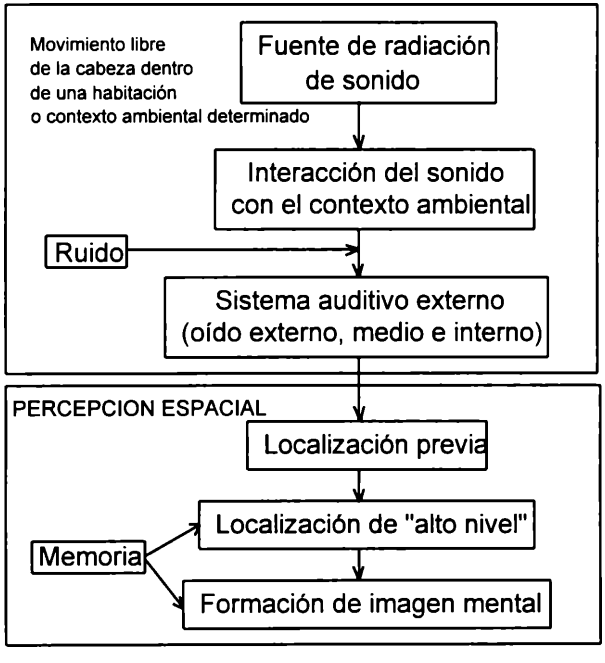


Fig. 2.3: Un modelo en bloques de la percepción espacial de una fuente de sonido

En este caso se dice que la fuente de sonido arriba como un *campo difuso* , debido a los efectos del *contexto ambiental*. Este es el principal componente del *medio* en el caso de la percepción espacial natural de un sonido. Otras fuentes de sonido dentro del contexto ambiental pueden ser consideradas indeseables bajo este punto de vista y ellas son clasificadas genéricamente como *ruido*.

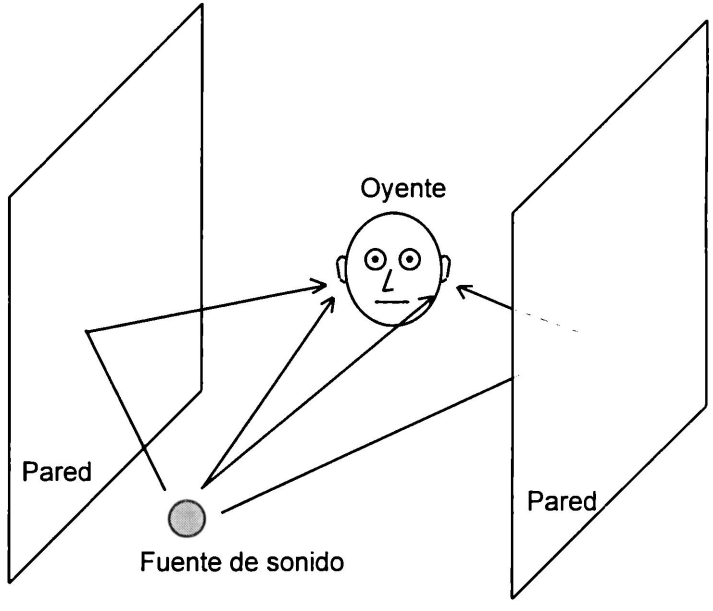


Fig. 2.4: Un modelo simplificado de percepción de una fuente sonora en un contexto cerrado. La señal llega por un camino directo y también por un camino indirecto producto de las reflexiones en las paredes e inclusive por las producidas por el piso y el techo (no mostrado aquí). La distribución de los caminos indirectos - reverberación- da información acerca del tamaño , clase del contexto ambiental y naturaleza de la fuente de sonido.

Ahora consideremos el modelo fuente-medio-receptor en el caso de un hipotético sistema de sonido 3D. Asumamos que el sonido va a ser reproducido a través de auriculares estereo. En este caso, un contexto ambiental esta ausente, y debe ser simulado en muchos casos. La fuente de sonidos previamente debe ser trasducida de la forma natural de energía acústica a una representación eléctrica por medio de micrófonos, u obtenida de algún medio de almacenamiento como un CD o disco magnético, o producida por un generador de tonos (sintetizador, *sampler*, etc.). Además, la cabeza del oyente no esta acoplada *a priori* al contexto ambiental de la fuente sonora. En lugar de ello, un par de transductores en forma de auriculares presentan el sonido contra el oído externo, tornandose ellos fuente y medio al mismo tiempo. El movimiento de la cabeza, utilizado normalmente para desambiguar la posición de cierto sonido,

debe ser tenido en cuenta, actualizando el sonido sintético en tiempo real. Si este efecto no se calcula, cuando el oyente gire su cabeza "arrastrará" todas las fuentes de sonido al mismo tiempo.

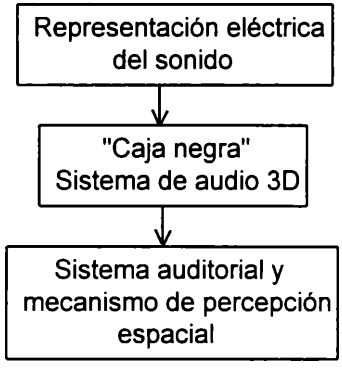


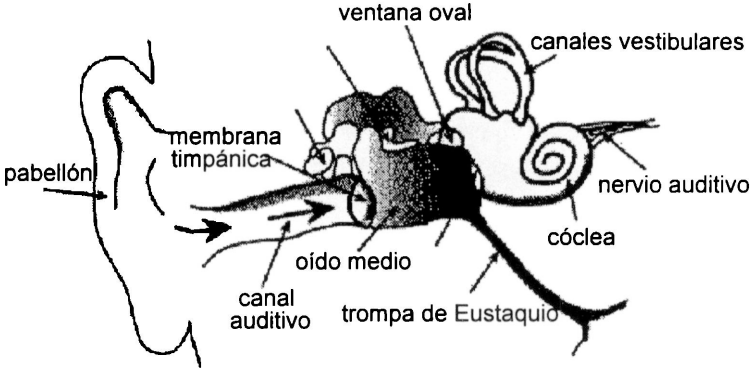
Fig. 2.5: Modelo de generación-percepción virtual espacial

La "caja negra" que comprende un procesador de sonido arbitrario transformará la representación eléctrica del sonido de alguna manera, para activar los mecanismos de percepción espacial del sonido y los aspectos cognitivos, determinando una imagen espacial de la fuente en el oyente. Esta asignación de posición puede ser similar a la vista en la figura 2.1 donde se representan los parámetros posicionales de una fuente natural de sonido. Pero hay que tener en cuenta que por medio de cierto dispositivo se puede recrear una sensación que no está presente en el mundo natural. Este contraste expresa una importante diferencia que aparece en los términos ambiente virtual y realidad virtual : el primer término implica una no-correlación con la realidad necesariamente, pues probablemente lo que queremos expresar es un fenómeno que no tiene contraparte en la realidad.

Desde el punto de vista del oyente, la percepción espacial natural y sintética consiste de las transformaciones físicas y perceptuales del frente de ondas del sonido, las cuales son ruteadas y preprocesadas por el sistema auditivo. La figura 2.6 muestra un esquema del sistema auditivo humano: el pabellón (porción visible del oído externo) transforma primariamente el sonido junto con las partes próximas del cuerpo, como la cabeza y los hombros. Esta transformación es de gran importancia para la localización espacial de una

fuente sonora. Sigue a esto el canal auditivo que da al oído medio, el cual consiste de la membrana timpanica y tímpano, compuesto por los huesecillos martillo, yunque y estribo. En el oído medio por medio de ellos la presión acústica es transformada a energía mecánica, la cual se convierte en una presión de fluido dentro del oído interno (la cóclea) producida por medio del movimiento de la ventana oval. La presión del fluido que depende del patrón de vibración externo hace vibrar la membrana basilar dentro del oído interno, la cual causa que numerosas fibras sensitivas se doblen. Esta acción activa un potencial eléctrico dentro de las neuronas del sistema auditivo, el cual es combinado a alto nivel cerebral con la información que proviene del otro oído. Estos procesos neurológicos son eventualmente transformados en una imagen percibida cognitivamente, la cual incluye atributos espaciales de la fuente.

Fig. 2.6: Sistema auditivo humano



Resúmen acerca de la percepción espacial del sonido

Azimuth y percepción de la elevación: pistas provenientes del tiempo y la intensidad

La pista más importante en la localización de la posición angular de una fuente de sonido involucra la diferencia que presenta el frente de ondas al arribar a nuestros oídos. Desde un punto de vista evolutivo, el posicionamiento horizontal de nuestros oído maximiza la diferencia de los eventos sonoros que se presentan a nuestro alrededor, permitiendo la audición de fuentes localizadas en la superficie del terreno fuera del campo visual. Veremos a continuación como la *diferencia interaural en tiempo* (ITD) y la *diferencia interaural de intensidad* (IID) influyen en la localización

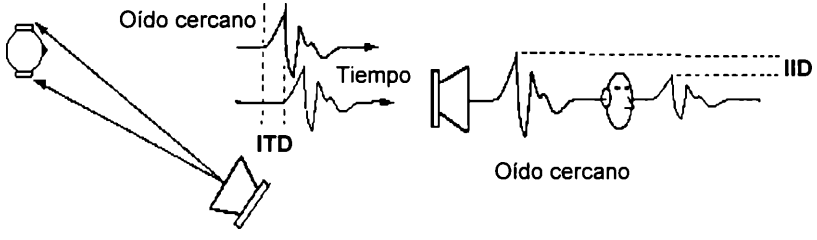


Fig. 2.7: ITD e IID se refieren a la diferencia en tiempo e intensidad que ocurre al presentarse una fuente sonora desde un costado del oyente.

La teoría básica que describe la localización a través de estas pistas se denomina "teoría duplex" y proviene desde principio de siglo [Rayleigh 07]. En esta teoría la proposición es que las IIDs son particularmente importantes en la localización de las altas frecuencias, mientras que las ITDs son importantes para las bajas frecuencias.

Los experimentos sicoacústicos realizados con auriculares que involucran la manipulacion de la ITD e IID, intentan investigar el fenómeno de *lateralización*. Este implica un caso especial de localización, donde se intenta parámetricamente mover una fuente de sonido para percibirlo fuera de la cabeza, generalmente en el eje interaural que intersecta los dos oídos.

Física de la lateralización

La lateralización proviene básicamente de un fenómeno que involucra cambios en la ITD e IID. Considere la siguiente situación hipotética: Suponga que existe un oyente con una cabeza perfectamente redonda y sin pabellón de la oreja, posicionado en una cámara anecoica a una cierta distancia de una fuente de

sonido de banda ancha, y con la fuente localizada en el plano medio horizontal de la cabeza. Modelar esta situación involucra calcular los dos caminos que presenta el frente de ondas desde el centro de la fuente hasta la entrada del canal auditivo.

Una simplificación adicional incluye poner estos canales en el medio de la esfera que representa la cabeza en el extremo del eje interaural ta como se ve en la figura 2.8. Con una fuente en la posición A a 0° de azimuth, el camino de llegada es igual, causando que el frente de ondas arribe a los tímpanos al mismo tiempo y con igual intensidad. En la posición B, la fuente de sonido esta aproximadamente a 60° de azimuth hacia la izquierda del oyente, haciendo que los caminos no sean iguales, esto causa que el frente de ondas llegue mas tarde al oído derecho con respecto al oído izquierdo.

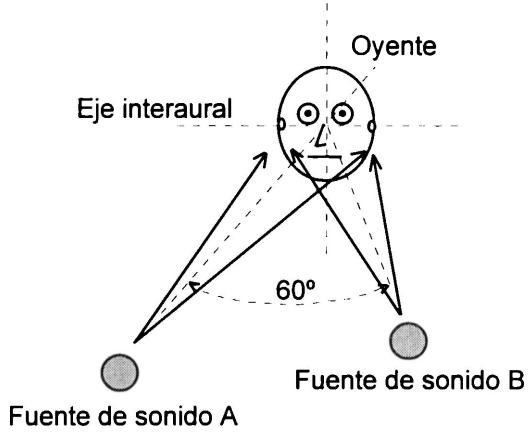


Fig. 2.8: Un oyente en una cámara anecoica con un fuente en frente suyo (A) y otra desplazada a izquierda unos 60° (B).

La diferencia en el largo del camino del sonido desde la fuente al oído es la base de la ITD (diferencia interaural en en tiempo), y el sistema auditivo es capaz de detectar los cambios de fases producidos por el retardo para frecuencias por debajo de 1 kHz.

La figura 2.9 muestra una cabeza esférica de radio *r*, con los oídos localizados en los extremos de eje interaural. Si un sonido distante incide sobre la cabeza

proviniendo desde la izquierda del eje de simetria XY a un angulo θ , para llegar al oído derecho el sonido debe recorrer mas distancia con respecto a lo que demanda para llegar al oído izquierdo, expresado esto por:

$$d = r \sin\theta + r\theta$$

La diferencia interaural t resulta:

$$t = \frac{d}{c} = \frac{r}{c}(\sin\theta + \theta)$$

donde c es la velocidad del sonido en el aire. La figura 2.10 muestra la ITD calculada con una velocidad del sonido en el aire de 344 ms^{-1} y un diametro de la cabeza de 175 mm.

La ITD es independendiente de la frecuencia de la fuente y de la forma de la cabeza. Experimentos muestran que la localización es más precisa para fuentes de sonido de banda ancha (por ej. ruido) que para tonos puros. Por ejemplo, la precisión de la localización para un click es del orden de los 8° mientras que para ruido blanco es de 5.6°. Los movimientos de la cabeza y los de la fuente de sonido son también pistas muy importantes en la localización binaural. El efecto más dominante es el cambio de la intensidad de la fuente: un desplazamiento de 15° puede resultar en un cambio de 2-3 dB [Kawalski 93].

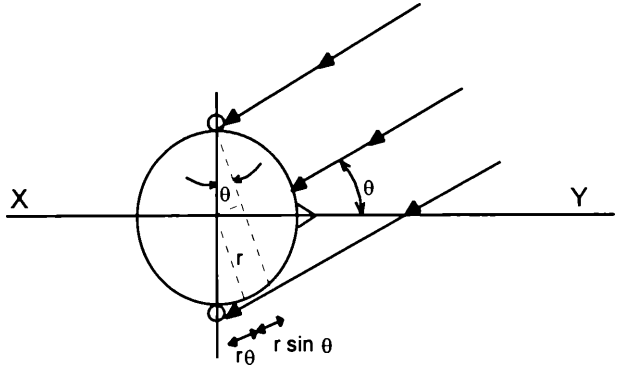


Fig. 2.9: Parámetros que influyen en la ITD.

Para frecuencias más altas de 1kHz, se produce un fenómeno en el cual un ciclo de la señal que arriba al oido mas lejano se solapa con el ciclo de la señal que arriba al oído más cercano, haciendo que otros factores comiencen a tener efecto en la localización.

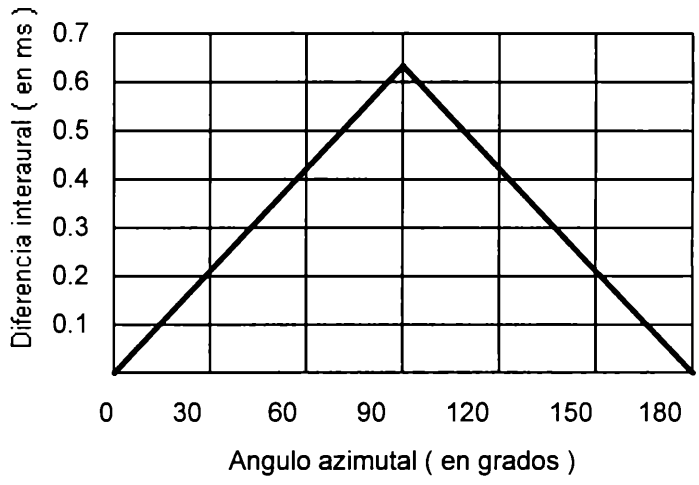


Fig. 2.10: Diferencia interaural calculada para una cabeza esférica.

Pistas ambiguas provistas por las ITDs e IIDs

Modificación espectral provista por el pabellón de la oreja

Mirando la figura 2.11, uno puede ver que dadas dos fuentes de sonido A y B producirán identicas ITDs e IIDs, lo mismo que las fuentes C y D. Esta afirmación es teórica pues la cabeza de una persona no es perfectamente esférica y los oídos presentan pabellón. Pero cuando las diferencias en las ITDs e IIDs son mínimas para diferentes fuentes sonoras que difieren notablemente en posición, una confusión potencial puede ocurrir si se usara solamente estas dos pistas (ITDs e IIDs).

En verdad, valores idénticos de ITD e IID pueden ser calculados para diferentes fuentes de sonido que pertenezca a la superficie cónica que se extiende desde el pabellón de la oreja hacia afuera, y pertenezca al plano perpendicular al eje interaural. A esta zona del espacio es denominada *cono de confusión*. Un modelo simplistico basado solo en ITDs e IIDs refleja un problema pues un oyente real es capaz de determinar las posiciones de las fuentes A,B,C y D, y allí las ITD e IID no porporcionan suficiente información para desambiguar la posición. De esta manera surge la hipotesis relativa al rol del filtrado espectral dependiente de la posición que proviene del oído externo.

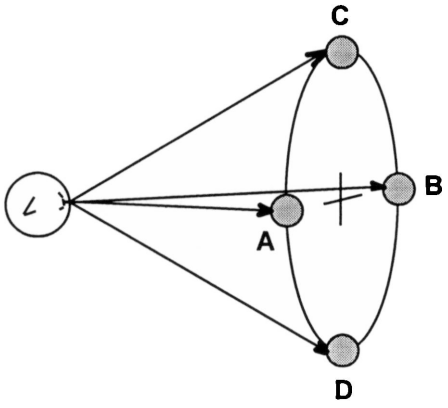


Fig. 2.11: Cono de confusión. Las fuentes A y B producirían entre ellas una confusión adelante-atrás y las fuentes C y D pueden originar una confusión de elevación.Inclusive A,B,C y D generan igual ITD e IID para el oyente.

Filtrado dependiente de la posición (HRTF filtering)

El filtrado espectral de una fuente de sonido antes de que llegue el frente de ondas al tímpano, es causado primariamente por el pabellón de la oreja y es conocido como HRTF (Head-Related Transfer Function). Las HRTF binaurales (terminologia para referirse a las HRTFs de los oídos derecho e izquierdo) pueden ser interpretadas como un filtrado dependiente de la frecuencia y de la posición de la fuente sonora, y en las cuales se incluye un retardo en el tiempo , como resultado del la forma compleja del pabellón de la oreja. Su forma asimétrica hace que se produzcan sutiles difracciones, resonancias y retardo de tal manera que pueden ser estos colectivamente trasladados a una única HRTF para cada posición del espacio. El uso de HRTFs es un componente clave en un sistema de sonido 3D. Esto es basado en la teoría que propone que el método mas preciso para producir un sonido espacializado con auriculares, es procesar el sonido original con estas HRTF, de tal manera que el sonido llegue al tímpano de la mismo manera como llegaría si la fuente fuera real en el espacio libre.

No solo el pabellón de la oreja afecta espectralmente la fuente de sonido, sino que la cabeza, los hombros e inclusive el torso imponen características específicas a las HRTF.

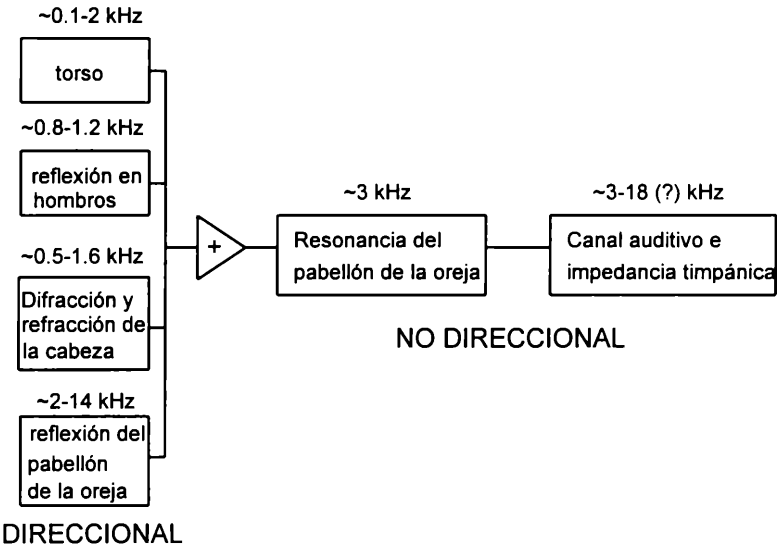


Fig. 2.12: Descripción de las componentes que conforman las HRTFs en función del aporte direccional que provee (según [Gierlich 92]). Es indicado también el rango de frecuencias afectado en cada etapa.

Localización usando HRTFs

El efecto particular que impone el oído externo al frente de ondas del sonido contribuye a la externalización y localización de una fuente sonora. Estas ideas sugieren que la localización perceptual de una fuente sonora usando auriculares puede ser posible si el filtrado espectral del pabellón también como las ITD e IID, son adecuadamente reproducidos en una fuente sonora dada. Existen muchos efectos acumulativos desde que el sonido es emitido hasta que llega al tímpano, pero todos ellos pueden ser expresados en una operación de filtrado única, en la cual se involucra atenuación espectral, retardo de tiempo , atenuación de intensidad y cambio de fase.

Medición de la HRTF

La naturaleza exacta de este filtrado puede ser medida con un simple experimento en el cual es producido un impulso muy corto en el tiempo por un parlante desde un posición específica. La transformación impuesta por los dos oídos es medida con unos pequeños micrófonos introducidos en el canal auditivo de una persona o maniquí con forma humana, lo mas cerca posible del tímpano. Esta situación es parecida a la que ocurre cuando golpeamos una campana con un martillo. El golpe de él (un impulso) transfiere energía en muchas frecuencias simultáneamente, las cuales son transferidas a la campana. El tono que es escuchado, depende de la característica física de la campana, y refleja que frecuencias son emitidas y cuales rechazadas o disminuídas.

Si la medición del impulso es realizada en los dos oídos simultáneamente, las respuestas obtenidas reflejarán también la diferencia interaural en el tiempo (ITD) y además la atenuación impuesta por la cabeza dependiendo de la posición relativa de la fuente.

Así la técnica de la medición a la respuesta al impulso permite medir todos los datos relevantes que influyen en la localización de una fuente sonora para una posición dada, un oyente en particular y un contexto ambiental determinado.

En la figura 2.13 vemos la representación equivalente de un pulso en el dominio del tiempo y en el dominio frecuencial (sólo se muestra amplitud). El paso de un dominio a otro se realiza por medio de la operación matemática denominada como transformada de Fourier..

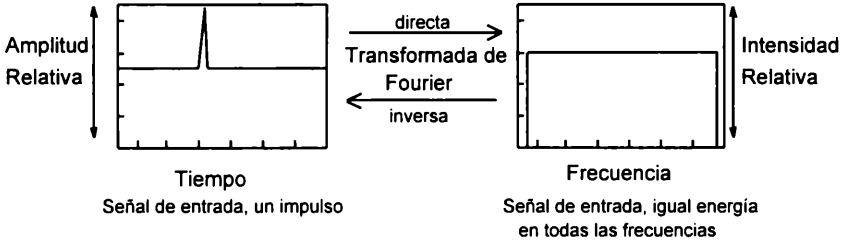


Fig. 2.13: Señal analítica original y sus representaciones equivalentes en el dominio del tiempo y frecuencial.

La figura 2.14 ilustra el efecto de la función de transferencia impuesta por el oído.En la figura 2.14 a) vemos lo que sucede cuando un impulso es enviado y transferido a los oídos a travez de un parlante localizado directamente enfrente del oído derecho del oyente. La interacción con el oído externo es medida con los micrófonos y es representada con linea continua para el oido izquierdo y con línea punteada para el oído derecho. Posteriormente en la figura 2.14 b) se ve la misma interacción pero representada en el dominio frecuencial. La diferencia de intensidad entre las mediciones del oído drecho e izquierdo es la IID. Los cambios de fase también aparecen en la medición, pero no fuero incluidos para mantener el diagrama más claro.

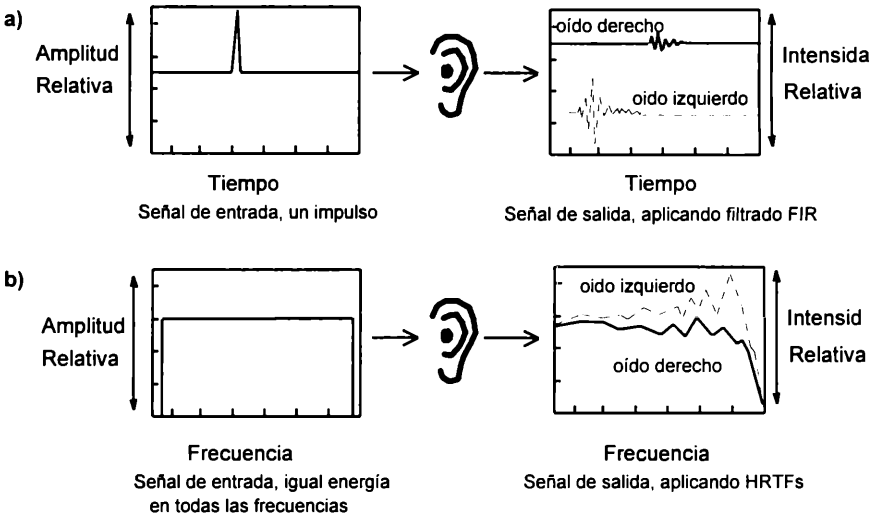


Fig. 2.14: Efectos del oído externo en el impulso analítico representados en el dominio del tiempo y en el dominio fecuencial.

Así cuando los filtros son construidos basandose en las características dependientes del oído, la representación en el dominio del tiempo es llamada filtro de respuesta finita al impulso o FIR y la representación en el dominio frecuencial es conocida como HRTF.

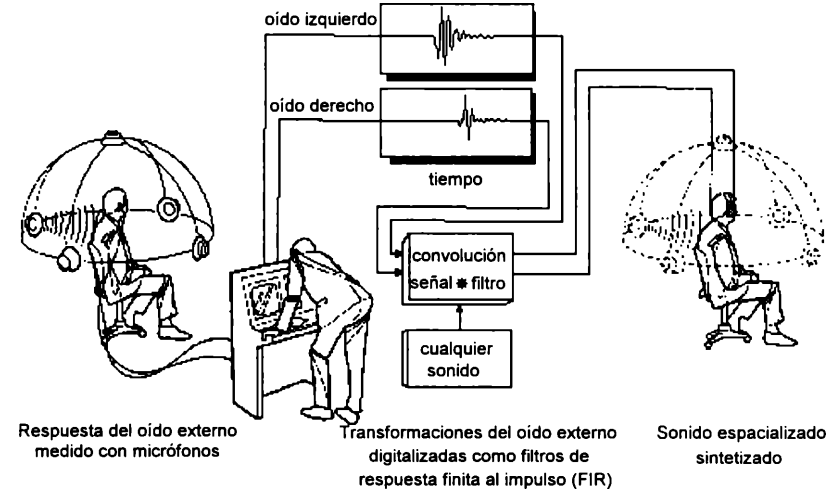


Fig. 2.15: Ilustración de la técnica de síntesis de fuentes de sonidos virtuales. Como ejemplo se mide un par de FIR para un pulso que es emitido en frente del oído izquierdo del oyente. La información obtenida es usada posteriormente para ser convolucionada con cualquier fuente de sonido monofónica y generar la sensación de escuchar a esa fuente desde la posición de donde se obtuvo los filtros. (Adaptado de [Wenzel 92]).

Filtrar en el dominio frecuencial es una operación de multiplicación punto a punto, pero en el dominio del tiempo la operación equivalente es la de

convolución. Filtrando un sonido arbitrario con estos filtros basados en HRTF , es posible imponer una característica espacial en una señal determinada, pareciendo finalmente que esta emana de una posición determinada. Por supuesto, la localización depende de otros factores tales como : contenido espectral de la fuente, HRTF utilizada para un oyente difiere de la HRTF medida de otra persona, etc. El filtrado realizado con HRTF no incrementa el ancho de banda de la señal original, solamente transforma las componentes frecuenciales que ya poseía.

Resúmen

Vimos cómo funciona y se puede generar el sonido 3D, el cual provee posibilidades interesantes no sólo en el campo de los entretenimientos o experiencias interactivas, sino que los sistemas que deben ser operados sin pistas visuales o con información de naturaleza espacial son altamente mejorados con este tipo de tecnología, tal como controladores de tráfico aéreo, pilotos de aeronaves o usuarios de sistemas de realidad virtual.

Usuarios no videntes son candidatos potenciales a la utilización de sistemas que incluyan este tipo de sonido espacializado. El estado del arte de la tecnología permite diseñar sistemas que generen sonido 3D, inclusive en tiempo real. En el caso de sistemas de información, esta tecnología es interesantísima como medio expresivo de ella, pues se permite explotar todo el espacio circundante como medio para representar información. El próximo tópico a estudiar es como un sistema digital puede procesar un sonido monofónico para generar un sonido 3D.

Procesamiento Digital de Señales y Sistemas

Capítulo

3

Resúmen

La presentación de información usando sonido y más precisamente sonido 3D, impone un estudio de la sicoacústica y modelización de la capacidad de localización de fuentes de sonidos que posee un ser humano. Si bien el mundo de los sonidos reales es analógico, la digitalización permite el procesamiento, almacenamiento y reproducción de manera electrónica y procesable computacionalmente. La digitalización es el primer paso para incorporar un sonido a un sistema digital. Este sonido un vez digitalizado, puede ser procesado por medio de diverso tipos de sistemas. Un tipo de ellos, son los sistemas lineales, causales e invariantes en el tiempo. Estos son muy importantes, fundamentalmente en este caso, pues el fenóneno de localización espacial de sonidos está ligado al efecto que impone el oído externo y la cabeza del oyente. Estos órganos imponen un procesamiento al frente de onda acústico tal como los sistemas anteriormente mencionados. Es por ello que resulta interesante estudiarlos, pues permitirá comprender como funciona la simulación sintética de sonido 3D.

Señales digitales

Una gran ventaja del procesamiento digital de señales o DSP (*Digital Signal Processing*), es que el procesamiento tiene lugar en el dominio digital, el cual ofrece flexibilidad y perfomance superior con respecto al procesamiento analógico, haciendo a las técnicas de DSP importantes con respecto al audio

digital. Las señales de audio que existen en la naturaleza son análogicas; por lo tanto las señales deben ser previamente digitalizadas. Este proceso es logrado por medio del muestreo o *sampling* de la señal en puntos discretos en el tiempo y asi cuantificando la amplitud de las muestras a un cierto valor de un conjunto permitido de valores . Como resultado, la señal original continua en el tiempo y en amplitud, $x(t)$, es reemplazada por una señal digitalizada, discreta en el tiempo y cuantizada en amplitud $x_i(nT)$, donde x_i es el valor de la amplitud en el instante $_i$, T es el período de sampling y n es un entero. Si el tiempo de sampling es normalizado, entonces en general la señal digitalizada es $x(n)$.

Por ejemplo, un señal señal senoidal de frecuencia f es representada en el dominio analógico como

$$x(t) = sin\omega\, t$$

donde la frecuencia angular es $\omega=2\pi f$. Si la señal es muestreada cada T segundos, la frecuencia de sampling es f_s Hz o w_s rad/seg. La relación entre el período de sampling y la frecuencia de sampling es $f_s=1/T$ y $\omega=2\pi/T$. La forma discreta de la ecuacion resulta

$$x(n) = sin\omega\, n$$

donde n es un entero. Las frecuencias de muestreo de 44.1 kHz y 48 kHz fueron adoptadas para manejar todo el rango dinámico del oído humano, que es algo menos que 20 kHz. Mas adelante veremos que la relación entre estos números no es arbitraria. Un ejemplo de señal continua y discreta se ve en la figura 3.1 y 3.2 . Los valores de n y t no son iguales en cada caso, la señal es muestreada a instantes discretos $t=nT$. Mirando los muestreos discretos es difícil determinar la frecuencia de la señal sinusoiudal en cuestión. Ella parecería ocupar ocho muestras por cada período. Asi se puede decir que la señal tiene una frecuencia que es un octavo del *sampling rate* o frecuencia de muestreo. La referencia en el dominio digital es la frecuencia de muestreo, todo es tomado relativa a ella. No es posible de discriminar visualmente la diferencia entre una senoide de 2 kHz muestreada a 32 kHz y una senoide de 3kHz muestreada a 48 kHz. Esto sucede pues en ambos casos la señal muestreada es un sexto de la frecuencia de muestreo. Por consiguiente, ambas mostrarán 16 *samples* o muestras por cada período de la senoide.

A primera vista parece que el proceso de muestreo introduce pérdida de información: la señal continua existe para todo instante pero solamente es muestreada a instantes discretos. Sin embargo, el muestreo a instantes discretos permite recuperar la señal original si el criterio de Nyquist es satisfecho. El teorema de Nyquist indica que la frecuencia de sampling debe ser al menos dos veces mayor que la frecuencia mas alta que se encuentre en la señal a muestrear. La violación de esta condición genera un fenómeno conocido como *aliasing* e impide recuperar la señal original además de generar señales inexistentes en la señal original. La medida de las amplitudes de cada instante muestreado require de una cuantificación de ese valor. La amplitud continua en el dominio analógico require un número infinito de bits para registrar adecuadamente el valor de la muestra. En realidad, solamente un número finito de bits es utilizado, por ejemplo, 16 para audio digital calidad CD.

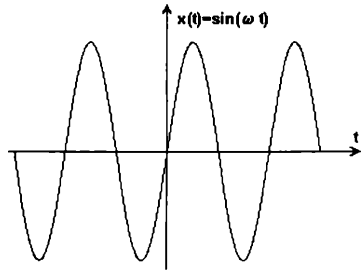


Fig. 3.1: Una señal continua en el tiempo esta definida para cada instante y posee valores continuos de amplitud

Volviendo a la figura 3.2, donde se muestra la representación de la señal de la figura digitalizada, la altura de cada barra representa el valor en ese instante . Esa altura puede tomar solo ciertos valores, mostrado por la línea de puntos, los cuales pueden ser mas altos o mas bajos que la señal original. Si la altura de cada barra es codificada con un número digital, entonces se dice que la señal está representada por el método de modulación de pulsos (o PCM, *pulse code modulation*). La diferencia entre la representación digitalmente cuantificada y el valor original analógico es denominado *quantization noise* (o ruido de cuantificación). Basicamente, este ruido está siempre presente en una señal digitalizada, pero cada bit extra que se introduzca para cuantificar la señal mejora la relación señal-ruido a razón de 6 dB por cada bit. Esto ocurre pues cada bit representa un factor de 2 , y 20*log 2 aproximadamente da 6 dB. Para sistemas de 8 bits el ruido es notable, percibiendose como un soplido de fondo. La tecnología de audio CD provee las señales con 16 bits de precisión, siendo este algo menor que el rango dinámico del oído, que es de alrededor de 120 dB (exigiendo 20 bits). Asi la cuantificación de la amplitud genera una distorsión, la cual puede ser particularmente notable para señales de poca amplitud, pues menos bits son utilizados para describir la señal.

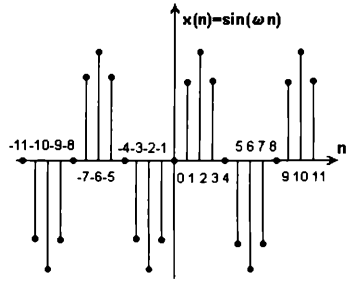


Fig. 3.2: Una señal discreta en el tiempo está definida solo en los instantes que se realizó el muestreo y puede sólo tomar ciertos valores de amplitud de un conjunto finito predefinido

La conversión del dominio análogico al digital se realiza por medio de dispositivos de hardware llamado conversores analógico-digitales (ADCs) y conj conversores digitales-análogicos (DACs) se realiza el proceso inverso.

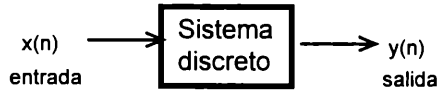
Sistemas lineales, invariantes y causales

Los sistemas discretos que son lineales, invariantes en el tiempo y causales son de gran importancia [Pohlman 91]. El sistema auditivo humano en lo referente a la localización de fuentes sonoras (oído externo) se comporta como tal. Un sistema discreto acepta una o mas entradas o *inputs* $x(n)$ y produce una o mas salidas o *output* $y(n)$ como se ve en la figura 3.3 a). En nuestro caso, la señal de audio que arriba a nuestros oídos es considerada la entrada, el oido externo y cabeza el sistema, y la señal que llega al tímpano la salida. Un sistema lineal posee dos propiedades- homogeneidad y superposición. Homogeneidad requiere que la amplitud de la salida sea proporcional a la amplitud de la entrada , para todo el rango de valores posibles de entrada. Si una entrada $x(n)$ produce una salida $y(n)$, entonces una entrada escalada $ax(n)$ producirá una salida escalada $ay(n)$ como se ve en la figura 3.3 b).

Superposición se refiere a la propiedad que cada señal de entrada será tratada independientemente de las otras. La entrada $x_1(n)+x_2(n)$ produce una salida $y_1(n)+y_2(n)$ como se ve en 3.3 c). Combinando estas dos propiedades de un sistema lineal, una entrada $a_1x_1(n)+a_2x_2(n)+...+a_Nx_N(n)$ producirá una salida $a_1y_1(n)+a_2y_2(n)+...+a_Ny_N(n)$. La entrada del sistema consiste en la suma de numerosas señales escaladas por los factores a_i . La salida es la suma de la respuesta del sistema a cada señal individual escalada adecuadamente segun fig 3.3 d). La linealidad es una propiedad importante pues significa que no existirá adición de nuevos componentes espectrales a la señal.

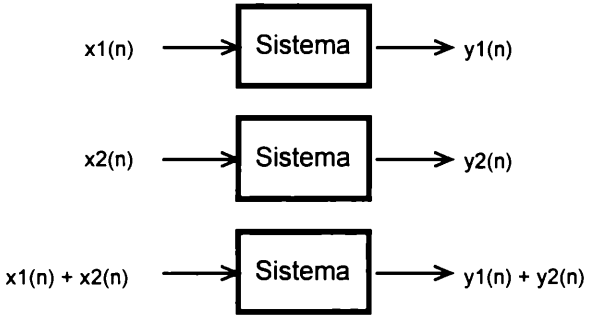
Un sistema invariante en el tiempo, mostrado en la figura 3.3 e), produce una salida *y(n-k)* para una entrada *x(n-k)*, para cualquier retardo o delay *k*.

Fig. 3.3: Propiedades de los sistemas discretos, lineales e invariantes en el tiempo.



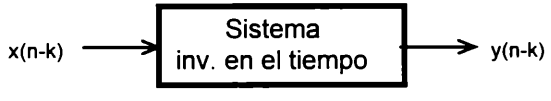
a) Relación entre la entrada y salida de un sistema discreto

b) Sistema que muestra homogeneidad



c) Propiedad de superposición

d) Un sistema lineal que muestra las propiedades de homogeneidad y superposición



e) Un sistema invariante en el tiempo

En otras palabras, retardar la entrada cierta cantidad de tiempo, retarda la salida en esa misma cantidad. Así, la salida es independiente del momento de orígen de la señal y sólo depende de la forma de la onda de entrada. Un sistema es

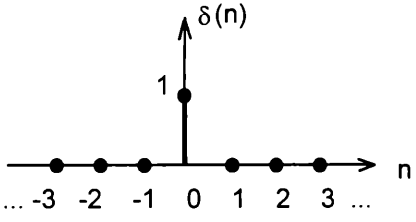
llamado causal o físicamente realizable si la salida no depende de valores futuros de la entrada.

La respuesta al impulso

La respuesta al impulso de un sistema lineal, denotado con *h(n)*, puede ser considerado al equivalente de su firma. Esta *h(n)* caracteriza completamente la respuesta del sistema. Matemáticamente, la respuesta del sistema a una función llamada delta o $\delta(n)$ es llamada respuesta al impulso. La función delta, como se ve en la figura , es un impulso en n=0 y es también llamada función de impulso. Ella contiene igual energía en todas las frecuencias y es definida en el dominio discreto como

$$\delta(n)=1, \text{ si } n=0 \quad \& \quad \delta(n)=0, \text{ si } n<>0$$

Fig. 3.4: La función delta o función de impulso

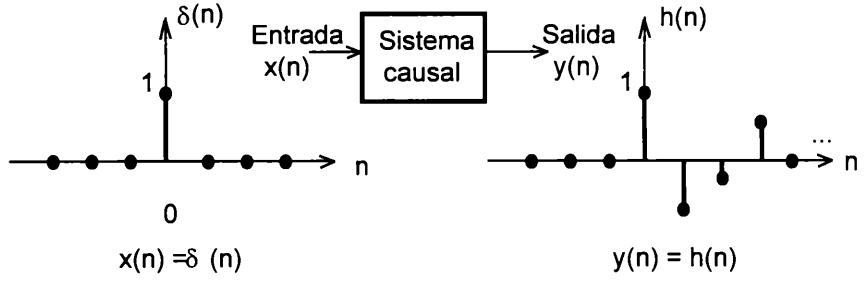


La respuesta al impulso puede ser usada para encontrar la respuesta en el dominio del tiempo, la respuesta en frecuencia y el cambio de fase característico del sistema. Un sistema se dice *estable* si cada entrada finita produce una salida finita (acotada en el tiempo) y si la respuesta al impulso posee energía finita. Una condición necesaria y suficiente para estabilidad referida a la respuesta al impulso de un sistema es

$$\sum_{n=-\infty}^{\infty} |h(n)| < \infty$$

Esto es, que la suma de los valores absolutos de cada muestra de la respuesta al impulso debe ser un número finito. Esto asegura que un una entrada finita producirá una salida finita.

Fig. 3.5: La respuesta al impulso de un sistema causal arbitrario



Si la entrada de un sistema causal es la función delta, entonces su salida será *h(n)*. La respuesta al impulso caracteriza al sistema; una arbitraria respuesta (de un sistema arbitrario) se ve en la figura . Debido a que la entrada ocurre solo a

n=0, *y(n)* deberá ser 0 para todos los valores negativos de *n*, pues *h(n)* no puede anticipar la respuesta hasta que la señal arriba a *n=0*. Como resultado, la causalidad implica que *h(n)=0* para *n<0*. Esta definición alternativa permite que *h(n)* y por lo tanto el sistema, pueda ser testeado directamente para determinar causalidad.

Convolución

La salida de un sistema lineal es la convolución de la entrada y la respuesta al impulso. La operación de convolución es denotada por el simbolo *. Matemáticamente, *y(n)* es la respuesta convolucionada de *x(n)* y *h(n)*.

$$y(n) = x(n)*h(n) = h(n)*x(n) = \sum_{k=-\infty}^{\infty} x(n)h(n-k) = \sum_{k=-\infty}^{\infty} h(k)x(n-k)$$

donde *x(k)* y *h(k)* representan las secuencias de entrada y respuesta al impulso, *x(-k)* y *h(-k)* representan las mismas secuencias invertidas, y *x(n-k)* y *h(n-k)* representan las secuencias invertidas corridas a derecha por *n* unidades. La expresión *x(k) h(n-k)* representa la multiplicación de la secuencia *x(k)* y la secuencia invertida y corrida *h(n-k)* , para un valor particular de *n* y *k*. La suma de todas estas multiplicaciones para un valor particular de *n* y todos los valores de *k* puede ser expresado como

$$y(n) = \sum_{k=-\infty}^{\infty} x(k)\,h(n-k)$$

Esto da el valor de *y(n)* para un valor particular de *n* , digamos n_1. Para encontrar el próximo valor de *y(n)*, $n=n_1+1$, la secuencia *h(n-k)* es corrida a la derecha una unidad mas. La multiplicación y suma es llevada a cabo como antes resultando *y(n)* para $n=n_1+1$. Una manera de ver la convolución es usar el hecho que el sistema es lineal e invariante en el tiempo. Por definición, la respuesta del sistema es la respuesta al impulso. La respuesta de un sistema lineal a un impulso escalado, es una versión escalada de la respuesta al impulso. La respuesta de un sistema invariante en el tiempo a un impulso retardado es una version retardada de la respuesta al impulso. La secuencia de entrada puede ser considerada como un tren de impulsos de amplitud variable, cada muestra arribando con un retardo determinado. Por lo tanto, cada muestra de la secuencia de entrada produce una version retardada y escalada de la respuesta al impulso, dependiendo de la amplitud de la muestra y de su posición temporal.

Estos resultados parciales son combinados para formar *y(n)* usando la propiedad de superposición. La salida en cualquier instante *y(n)* es la suma de las partes de la respuesta al impulso producidas por las entradas corridas y escaladas para un instante de tiempo determinado.

Espectro

Existen dos dominios o maneras de representar señales que son generalemente usadas: el dominio del tiempo y el dominio de la frecuencias. En el dominio del tiempo, se estudia la manera en que la señal cambia a lo largo del tiempo. En el dominio frecuencial, se estudia la ausencia o presencia de ciertas componenes frecuenciales. La musica y el sonido de fuentes naturales consisten de diferentes frecuencias de amplitudes variables, y si ellas son pasadas a traves de un banco de filtros , la salida denotará la presencia y ausencia de ciertas frecuencias y su amplitud relativa. Esta información es la que se observa en un analizador de espectro, donde la señal de entrada es particionada en sus componentes frecuenciales. En otras palabras, esta presentación es una visión de la señal en el dominio frecuencial. Intuitivamente, si la variación en el dominio del tiempo es lenta, hay indicios de la presencia de bajas frecuencias. Un cambio rápido en la amplitud significa que hay presentes altas frecuencias en la señal analizada. Aplicando la transformada de Fourier a una señal de audio, uno puede determinar la amplitud de las frecuencias responsables de una forma de onda particular.

Es fácil de encontrar el espectro en tonos simples con una frecuencia definida y con un régimen periódico de repetición. En los sonidos cotidianos y música, el espectro cambia rápidamente con el paso de tiempo. Si la forma de onda es estudiada por una largo período, entonces resultará en una medición mas precisa del espectro. El espectro obtenido en este caso, sin embargo es el valor del período completo de muestra, no brindando informacion del espectro en un subintervalo de tiempo determinado. La precisión de la medición del espectro (en términos de resolución frecuencial) disminuye si es medido en un intervalo menor de tiempo. El problema puede ser mejor entendido si en el caso extremo el subintervalo es de solo una muestra. Una sola muestra no da información acerca del espectro de la señal analizada. Como resultado, la medida del espectro es un compromiso entre la precisión del espectro obtenido y la informacion espectral (granularidad de la informacion) durante un intervalo en particular.

Transformadas

Las transformadas son herramientas matemáticas que permiten mover la información del dominio del tiempo a un dominio frecuencial y viceversa. Existen una gran cantidad de transformadas. Las transformadas pueden ser categorizadas básicamente en transformadas continuas, transforamadas de serie y transformadas discretas. Las transformadas continuas son aplicadas a señales continuas en el tiempo y frecuencia. Si una señal continua en el tiempo posee solo ciertas componentes frecuenciales, entonces puede decirse que la señal posee espectro discreto o de líneas. En este caso, la transformada continua se reduce a una transformada de serie. Así las transformadas de serie son un caso especial de las transformadas continuas y son aplicadas a señales continuas en el tiempo, con componentes frecuenciales discretas. Las transformadas discretas son aplicadas a señales que son discretas en el tiempo y en las frecuencias, tal como son las que aparecen en los sistemas digitales.

La transformada de Fourier

La transformada de Fourier es un caso especial de la transformada de Laplace. Es una operacion matemática que mapea una función en el dominio del tiempo *x(t)* a una función del dominio frecuencial *X(jw)*. Esto es

$$X(j\omega) = \int_{-\infty}^{\infty} x(t)\, e^{-j\omega t}\, dt$$

La transformada inversa de Fourier mapea una funcion del dominio frecuencial *X(jw)* a un función del dominio del tiempo *x(t)* así

$$X(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} x(t)\, e^{j\omega t}\, d\omega$$

Físicamente , *X(jw)* describe el espectro de la señal *x(t)*. Sin embargo, si *X(jω)* es el espectro de un sistema, entonces este describe la salida espectral de un sistema cuando la entrada es una función de impulso. En otras palabras, *X(jω)* es la respuesta frecuencial del sistema en cusetión. La transformada de Fourier efectivamente descompone la señal del dominio del tiempo en sus componentes frecuenciales constitutivos.

La transformada discreta de Fourier

La transformada discreta de Fourier (DFT) es un ejemplo de una transformada discreta. La DFT es la transformada de Fourier de una señal muestrada *x(n)*. Si la señal *x(t)* es uniformemente sampleada y convertida a *x(n)*, la integral que define la transformada de Fourier puede ser aproximada por la suma

$$X(j\omega) \cong T \sum_{n=-\infty}^{\infty} x(nT)\, e^{-j\omega nT}$$

donde T es el período de muestreo. Esto proporciona valores uniformemente espaciados de frecuencia ω. En situaciones prácticas debido a limitaciones de hardware, solamente un numero finito de muestras es considerado para el análisis. Esto es llamdo una transformada de Fourier finita. Si la muestra consiste de N puntos, la DFT de N puntos y su inversa es definida respectivamente como

$$X(m) = \sum_{n=0}^{N-1} x(n) W_N^{nm},$$

$$x(n) = \frac{1}{N} \sum_{m=0}^{N-1} X(m) W_N^{-nm}$$

donde $W_N = e^{-j2\pi/N}$. Debe tenerse en cuenta que W_n^k (=exp[-j2πk/N]) es igual a $W_N^{k+\pi N}$, pues

$$\begin{aligned} W_N^{k+pN} &= \exp\,[-j2\pi\,(k + \pi N) / N] \\ &= \exp\,(-j2\pi\, k/N)\, \exp(-j2\,\pi\pi) \\ &= \exp\,(-j2\pi\, k/N) * 1 \\ &= \exp\,(-j2\pi\, k/N) = W_n^k \end{aligned}$$

En los algoritmos de DFT, *X(m)* es usualmente computado como *N* componentes frecuenciales igualmente espaciadas en el intervalo 0 a w_s. Los componentes espectrales estan separados por w_s/N. La DFT es evaluada a frecuencias discretas dado por $w=m(w_s/N)$, para m=0,1,2,...,N-1. El término *X(m)* representa el nivel o amplitud de la frecuencia $m(w_s/N)$ contenida en la señal *x(n)* y es frecuentemente referida como componente M-ésima. Estas son las únicas señales que pueden componer a *x(n)*, pues estas son las únicas frecuencias cuyos períodos estan enteramente relacionados a N. Las frecuencias presentes son todas armonicamente relacionadas a la fecuencia fundamental w_s/N (correspondiente a $m=1$), excepto para la componente de continua o DC (

correspondiente a $m=0$). Asi la DFT descompone la señal en el dominio del tiempo en término de un conjunto de frecuencias discretas armonicamente relacionadas, es decir por relación entera. El número de componente frecuenciales es usado para especificar una armonica particular de la frecuencia fundamental. La amplitud de cada componente da una medida de la potencia del espectro, debido a que la potencia de esa componente esta relacionada con el cuadrado de su amplitud.

Es sabido que para una señal real (todas las seniales de audio son asi) , la DFT es par en magnitud. Ademas es periódica. Como resultado de esas dos propiedades, solamente la mitad positiva de las frecuencias son mostradas de las DFT's . Debido a la presencia de frecuencias positivas y negativas (excepto para la componente de continua y la frecuencia de Nyquist), la amplitud de la respuesta de cada componente es la mitad de lo que debería ser. Esto es debido a que la energía de estas frecuencias es igualmente dividida entre las frecuencias positivas y negativas. También es sabido que la convolución en el dominio del tiempo es equivalente a la multiplicación punto a punto en el dominio frecuencial y viceversa. Esta propiedad es llamada de dualidad [Oppenheim 89]

En la evaluacion de una DFT cada $X(m)$ requiere N multiplicaciones complejas y $N-1$ adiciones complejas. Debido a que existen N componentes espectrales, la DFT requiere aproximadamente N^2 multiplicaciones complejas y $N(N-1)$ adiciones complejas. La memoria requerida es $2N$ para $x(n)$ y $X(m)$,y N^2 para los W_N^{nm} coeficientes. La implementacion de una DFT en hardware muestra una complejidad computacional proporcional a N^2 en computación y memoria. Por ejemlo, una DFT de 1024 puntos requiere cerca de 1 millon de multiplicaciones y adiciones complejas. Así, una DFT es típicamente un proceso que no se lleva a cabo en tiempo real

La transformada de Fourier rápida (FFT)

La transformada rápida de Fourier o FFT (Fast Fourier Transform) se refiere a la colección de algoritmos que explotan la simetría y periodicidad de los coeficientes W_N^{nm} para acelerar el calculo de la DFT. El algoritmo de la FFT requiere que N sea un entero potencia de dos. Para secuencias arbitrariamente largas, se completa con ceros para llegar a la próxima potencia de dos. El calculo de la FFT requiere $N \log_2 N$ pasos de cómputo y un tamaño de memoria de $2N$. Numericamente, la DFT y la FFT dan el mismo resultado, sin embargo la

FFT computa el resultado mucho mas rápido usando una configuración especial de coeficientes. Para N=1024 la cantidad de computaciones estan en el orden de las 10000. Esto implica un factor de aceleracion de 100 con respecto a la DFT. Esta mejora en velocidad frecuentemente permite el procesamiento en tiempo real. La FFT puede ser implementada por software o por hardware específicamente diseñado.

Conclusión

La naturaleza maneja variables analógicas que para ser procesables digitalmente necesitan ser cuantificadas con valores finitos de precisión en tiempo y amplitud. Esta digitalización permite recrear posteriormente con un proceso de conversión digital-analógica, fenómenos tales como sonidos previamente capturados. Algunos sistemas, tal como el oído externo humano y el medio de transmisión pueden considerarse sistemas lineales, invariantes en el tiempo y causales. Estudiando sus propiedades podemos saber como se puede procesar una señal arbitraria para darle la característica que le impone un cierto sistema.

Si basicamente a una señal de audio la procesamos con el efecto del oído externo (el sistema lineal en cuestión), podremos recrear una sensación espacial dependiente del procesamiento realizado. Si ese procesamiento se realiza convolucionando los filtros adecuados (HRTF's) a un sonido monofónico arbitrario, podremos regenerar ese sonido con los atributos espaciales deseados, a pesar de estar escuchándolo con auriculares.

Hipermedia

Capítulo

4

Resúmen

Hipermedia se refiere a un conjunto de tecnologías que tratan con una manera de organizar, proveer asociaciones entre piezas relacionadas de información, y mostrarlas adecuadamente. Aunque muchos sistemas que se denominan hipermediales han sido desarrollados, el concepto de hipermedia es aún constantemente definido y revisado.

Sin entrar en detalle en el aspecto formal o descripciones de sistemas de autoría, revisaremos los items globales mas importantes a tener en cuenta en un sistema de hipermedia, para compararlos y ver de que manera se pueden solucionar o implementar ellos en el sistema HiperAudio. A medida que se va comentando cada tópico, es útil pensar como se puede implementar este sin ningún tipo de pista visual.

Por qué hipermedia?

Hipermedia es una tecnología que trata de resolver el problema de la organización y acceso a información de naturaleza multimedial. En lugar de organizar información como un flujo secuencial de datos, hipermedia codifica la información en pequeñas unidades autocontenidas llamadas *nodos*. Los nodos que estan relacionados entre sí, mantienen una relación gestionada por el sistema. La entidad conceptual que mantiene esa relación es denominada genericamente *link*, la cual genera una red de información tal como se ve en la figura 4.1.

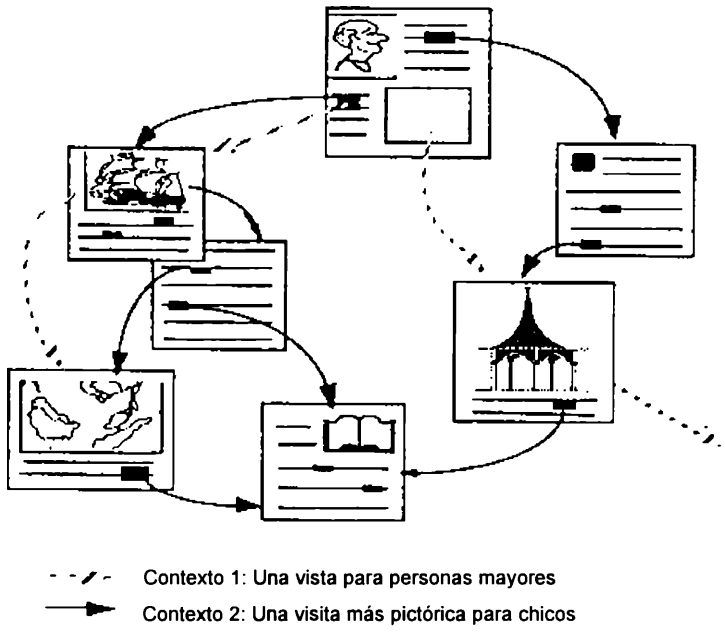


Fig 4.1 : En una aplicación hipermedial la información está contenida en unidades denominadas nodos. Por medio de una acción implícita o explícita el usuario puede activar links y así navegar a otros nodos. En este caso, diferentes tipos de links pueden esta activos dependiendo del contexto seleccionado por el lector.

En lugar de acceder a la información buscando a través de índices o pasando páginas, el usuario de un sistema tal como el propuesto, accede a la información seleccionando links a través de un cierto camino de información asociada. Estos links procesables por la máquina subyacente son los que dan la potencia al modelo hipermedial. El tipo de información que puede ser manejado por un sistema de hipermedia es diverso: se puede incluir desde medios estáticos tal como texto, imágenes y gráficos; medios dinámicos tal como sonido, video y animación; y medios variables tal como una planilla de cálculo.

Hipermedia en su forma básica, representa una idea particular en la forma de acceso a la información y su exploración. Presenta un potencial para una forma de exploración adecuada para gran cantidad de información multimedial al ritmo que precisa el usuario y de acuerdo al interes de él. Por lo tanto hipermedia se ajusta bien con las tendencias de dar mas potencia al usuario que interactua con el sistema. Debido a esto, hipermedia es una herramienta ideal para aprendizaje y simulación.

El potencial completo de hipermedia puede ser interpretado completamente comparándolo con los usos tradicionales de las computadoras y las posibilidades

de los medios impresos tradicionales. Entre las posibilidades interesantes que hipermedia provee, tenemos:

• Permite al usuario interactuar con el sistema para ver o componer información en una manera no-lineal y asociativa

• Permite una presentacion dinámica de la información con una dualidad: que es presentado y como es presentado

• La información es "activa", asi el usuario interactua directamente con la información para realizar operaciones tal como navegar a traves de un link o activar un programa asociado

• Provee una base consistente para la asociación y visualización de grandes cantidades de información heterogénea

Son estos hechos los que permitieron a los primeros investigadores imaginar la creación de gigantescos sistemas de información, los cuales relacionarían toda la información mundial. Esta idea actualmente es parcialmente implementada a traves de la WWW (World Wide Web).

Características de los sistemas actuales

A pesar de los diferentes tipos de aplicaciones de sistemas hipermediales y el énfasis particular de cada uno de ellos, un número de características en común pueden ser identificados en muchos de los sistemas. Estas características en común son discutidas en términos de: creación de la información, organización, presentación, interacción y obtención.

• Creación de la información y organización: Muchos sistemas organizan la información basandose en el modelo de nodos y links, donde la entidades básicas son un conjunto de nodos interconectados por un conjunto de links. A través de un editor de hipermedia, el usuario puede crear nuevos nodos y adicionar links a cualquier nodo. El usuario tambien puede alterar el contenido de los nodos o los atributos del link. Los links son generalmente generados y mantenidos manualmente por los usuarios.

• Interacción y presentación de la información: usualmente los sistemas utilizan la metáfora de libro como método de presentación a los usuarios. Como en un libro, el punto principal de acceso a estos sistemas es la tabla de contenidos o índice. Ciertas funcionalidades del sistema son provistas para permitir a los usuarios realizar marcas o bookmarks en nodos importantes a revisitar, remarcar

ciertos pasajes, tomar notas que se pueden incluir junto con los datos originales, y mantener una historia o lista de nodos recientemente visitados para propósito de backtracking o revisión de información obtenida recientemente. Generalmente se provee herramientas para incluir gráficos, sonido y video. Actualmente los sistemas presentan interfaces multiventanas y algunos proveen la posibilidad de representación de la red de nodos para ayudar la navegación.

• Recuperación de información: En términos de de recuperacion de información, muchos sistemas enfatizan en el browsing o inspección exploratoria de la información sobre el modelo de nodos y links. El browsing puede ser llevado a cabo de tres maneras:

- Siguendo los links entre los nodos
- Seleccionando los nodos apropiados de un cierto grafo presentado
- Examinando los nodos almacenados en la lista de historia y bookmarks

La información tambien puede ser accedida buscando en la red por medio de un método de interrogación o *query*, por medio de una combinación lógica de strings, palabras clave o valores de los atributos de las entidades.

Limitaciones de los sistemas actuales

La generación actual de herramientas de autoría de aplicaciones hipermediales es basado en el modelo de nodos y links. Aunque este modelo es muy simple, flexible y eficiente, es de muy bajo nivel para soportar integración de información a gran escala. En particular, la generación actual de sistemas de hipermedia muestra una falta de soporte en lo siguiente puntos:

• Orientación del usuario dentro de la red de información: el problema de la desorientación ocurre notablemente cuando la red excede los 1000 nodos [Conklin 87]. El usuario no sabe donde está, como llegó o como puede hacer para alcanzar lo que precisa desde donde está.

• Soporte de un alto nivel de abstracción en la información: un alto nivel de abstracción provee una manera de representar un grupo de nodos y links como una única entidad con una sintaxis bien definida y propiedades semánticas claras. Niveles de abstracción sucesivos posiblemente reduzcan una vasta cantidad de nodos, en un pequeño número de ellos, con un concepto bien

definido. Con mejor manejo, estos nodos de alto nivel pueden ser usados para proveer un contexto a los usuarios durante el browsing, aliviando el problema de la desorientación.

• Facilidades para ayudar a los usuarios a organizar y clasificar la información de una mejor manera: los sistemas actuales delegan al autor la creación y mantenimiento de links entre nodos. El proceso manual de creación de links es uno de los esfuerzos principales en la autoría [Glushko 89] . Inconsistencias surgidas de la creación cooperativa de documentos es un aspecto a tener en cuenta. Por ej.: en WWW es factible encontrar links que apuntan a hosts o documentos inexistentes actualmente o que fueron movidos

• Presentación inteligente de la información y filtrado: esto requiere una habilidad del sistema para decidir no solo que y cuanta información será presentada al usuario, sino también como debe ser presentada. Los sistemas actuales no proveen la capacidad deductiva necesaria como para soportar estas características.

• Interacción y acceso consistente a la información multimedial: un atributo esencial de la hipermedia, es que la información es "activa", esto es, que el usuario puede interactuar directamente "clickeando" sobre la región de interés, para realizar cierta operación que implica obtener mas información acerca del tópico en cuestión. Casi todos los sistemas existentes, donde el paradigma de presentación de medios es 2D, tal como ocurre con el texto, imágenes y gráficos, las ideas precedenentes funcionan. Sonido, video y gráficos 3D son en general introducidos como medios pasivos para suplementar la información que aportan los medios estáticos. Usualmente no se proveen operaciones de búsqueda sobre estos medios.

Tópicos de interés en la investigación de sistemas hipermediales

Como solución a las limitaciones anteriormente citadas, muchas ideas han sido propuestas en el área de construcción de sistemas y usabilidad de ellos. Brevemente indicaremos cinco áreas a tener en cuenta en la próxima generación de sistemas de hipermedia:

• Uso de semántica y estructuras de alto nivel para proveer mejor browsing: por ej: se han propuesto mecanismos de frames para extraer, organizar y presentar un subconjunto de información basado en el interes del usuario; o por ejemplo jerarquías de objetos representados tridimensionalmente que dan una idea de la topología de la red y la relación entre las entidades.

• Integración de herramientas de búsqueda y *query*: muchos de los sistemas de hipermedia basan la búsqueda en el matching exacto de palabras clave, pero este no llega a ser satisfactorio [Crouch 89]. Facilidades de búsqueda basadas en contenido y estructura son útiles herramientas . Existen sistemas que buscan por contenido imágenes e inclusive se plantean métodos para buscar sonidos por contenido [Hirata 93].

• Desarrollo de sistemas dinámicos de hipermedia: actualmente los sistemas de hipermedia reflejan una naturaleza estática de la información con respecto a la creación de asociaciones en *run-time* o tiempo de corrida. Una solución a este problema es la creación dinámica de links. Un link dinámico es definido como uno, en el cual la asociación que genera entre nodos es determinado en tiempo de corrida , basándose en el contexto y contenido de los nodos. Técnicas automáticas basadas en motores de asociación basados en conocimiento se han propuesto [Belkin 87].

• Desarrollo de interfaces de usuario (UI) mas efectivas: el desarrollo de herramientas de ayuda para la navegación dentro del contexto de nodos y links ha sido intentado por muchos autores. Las herramientas de navegación intentan presentar un *overview* o resúmen de la red hipermedial a los usuarios. Variaciones de estas herramientas han sido desarrolladas desde browser gráficos 2D [Halasz 87], mapas [Feiner 88] o herramientas navegacionales 3D [Fairchild 88]. Aunque estas herramientas son útiles para la navegación exploratoria, ellas se han mostrado ineficientes para grandes hiperbases [Monk 89]. Para visualizar grandes hiperbases, técnicas de abstracción de grafos han sido propuestas , tal como *fisheye views* o clustering de nodos. El uso de metáforas físicas para proveer una interfaz consistente de tal manera que el usuario interactue en un ambiente familiar, es una política que esta ganando popularidad. Algunas de las metáforas mas comunes son: libros, tarjetas, bibliotecas o tours guiados. Sin embargo, una sola metáfora aparece insuficiente para presentar las diferentes funcionalidades y vistas de una hiperbase. Un sistema hipermedial con una

interfaz extendible y el soporte de una variedad de metáforas parece ser esencial para el futuro de aplicaciones hipermediales.

• Integracion homogénea de información multimedial en un ambiente hipermedial: hipermedia presenta naturalidad para la manipulación directa de información en la pantalla. Este paradigma funciona bien para texto y gráficos debido a que su contenido esta codificado en unidades discretas y manipulables como frases o entidades gráficas. Este no es el caso de medios tales como el sonido o video. Estos medios presentan nuevos tópicos con respecto al manejo de links temporales e interacción en una situación dinámica. Por ej: íconos dinámicos en una presentación de video permite al usuario navegar dentro de un clip de video durante su reproducción [Brondmo 89]. La inclusión de gráficos 3D, animación y dispositivos de reproduccion inmersivos tal como HMDs (*Head Mounted Display*) dan interesantes posibilidades para la manipulación y simulación de mundos virtuales, los cuales serían contenedores de información navegables con una metáfora espacial .

Resúmen

La modalidad de acceso hipermedial provee amplias ventajas para el acceso exploratorio de información, pero también deja temas abiertos tal como los planteados previamente. Mas aún, para interfaces no visuales se deben proveer métodos y modalidades para resover esa problemática. Desafortunadamente aún no existen los guidelines necesarios para atacar esos desafíos.

Con los resultados provistos por el uso del sistema propuesto probablemente se establecerá cierto mapping de las facilidades existentes en los sistemas que corren sobre una GUI.

Introducción a los Displays Acústicos

Capítulo

5

Resúmen

El estudio de los sonidos desde el punto de vista físico o sicoacústico solo refleja una parte en el diseño de un display acústico. En este capítulo se revisarán las ideas básicas de la representación auditiva de sonidos desde el punto de vista cognitivo, es decir: interpretar los sonidos no en función de sus características físicas sino de la manera en que estos son asimilados y como contribuyen a la formación de imágenes mentales.

La idea fundamental de un display acústico es la de poder representar cierto contenido de información a través de sonidos. Esta información puede referirse a orientación en un cierto ambiente, el estado de un sistema o inclusive la representación acústica de un conjunto de datos (sonificación).

La manipulación de las variables que definien un sonido, permiten mapear cada una de ellas a diferentes parámetros del sistema en función de la aplicación del display acústico. Las ventajas y desventajas de esta modalidad de percepción de un sistema también es comentado.

Displays Acústicos: que son

La investigación acerca de los displays acústicos se refiere a la forma en que usamos sonidos en la interacción Hombre-Máquina para representar cierta información. La función de un display acústico (DA) es ayudar al usuario a monitorear y comprender cierto sistema a través del sonido percibido.

Una extensión a este concepto surge en la idea de pantalla virtual acústica 3D [Wenzel 92], la cual puede ser definida como un medio preciso para transferir información a un ser humano usando modalidades auditoriales, las cuales combinan caracteristicas direccionales y semánticas de objetos dinámicos, los cuales representan entidades de un ambiente simulado.

Si la codificación sonora expresada por el sistema se presenta a través de voz, entonces el DA está explotando el repertorio del lenguaje y su significado cognitivo. Si la codificación está expresada a través de sonidos *nonspeech*, el DA explotará el conocimiento evolucionario adquirido a través del uso del sistema y las pistas cognitivas que generan la naturaleza de las señales percibidas.

Sonidos en la vida real

El sonido es una crucial fuente de información acerca nuestro ambiente físico. Las vibraciones de aire que llegan a nuestros tímpanos son interpretadas como sonidos, y ellas contienen la información acerca la naturaleza mecánica de la fuente de esas vibraciones, el espacio que ocupa, la dirección y la posibilidad de percibir la distancia de la fuente.

Estas simples acciones son las que entran en juego en el diseño de un DA. Un tema importante a tener en cuenta, es que un DA es una forma de convenir información sin pistas visuales. Tomemos este ejemplo de la vida real para ver la importancia del oído en la percepción espacial: si estamos en la calle y escuchamos que viene un camión velozmente hacia nosotros por detrás, podemos: separar el ruido del camión de otros ruidos del tránsito (*stream segregation*), determinar cuan lejos está (distancia) y su posición (direccionalidad). Estas características hacen que nuestra percepción auditiva posea una característica omnidireccional, pues podemos percibir lo que *a priori* no se puede ver. Así, la visión provee enorme detalle en una dirección determinada, pero el sentido auditivo nos permite monitorear todas las direcciones simultáneamente. Este es un importante tópico a explotar en el diseño de un DA. De la misma manera en la cual nosostros percibimos texturas de los objetos que observamos, también percibimos texturas a través del oído. Mas que proveer información acerca del color y calidad de la superficie, la textura sonora nos provee una sutil información acerca del proceso físico que estamos percibiendo. Por ejemplo podemos saber si un eje giratorio esta girando con aceite suavemente o sin el con crujidos y chirridos.

La capacidad de atender a un grupo de sonidos y asociarlos a una sola entidad conceptual se denomina *stream segregation* [Bregman 94]. Las forma en que trabajan estas habilidades de alto grado cognitivo no son completamente comprendidas actualmente. Esta capacidad de stream segregation es la que permite aislar el camión en el ejemplo anterior de la calle o la melodía que ofrecen los vientos en una orquesta sinfónica. Este hecho permite que en un DA podamos explotar este fenómeno, atendiendo selectivamente a una sola fuente de información auditiva, pero permitiendo simultáneamente la percepción simultánea de otras fuentes en paralelo.

Beneficios de un Display Acústico

Los beneficios que surgen de la aplicación de estos conceptos para la presentación de información a través de un DA para un ciego es evidente, pues de alguna manera se puede convenir protocolos y formalismos para interactuar con ciertas entidades. En la siguiente tabla se resume en general, cuales son las características positivas de un DA.

Ventaja	Aplicación
Ojos libres	Habilidad de monitorear otras variables que no se perciben actualmente con la vista, visualizacions complejas o dinámicamente cambiantes, interfases para impedidos visuales
Rápida detección y alerta	Monitoreo, ambientes de alto stress (por ej. controladores aéreos, monitoreo médico, etc.)
Orientación	Exploración de datos (por ej. para indicar áreas de interés), obtener indicación a donde mirar
Backgrounding	Monitoreo o exploración de conjuntos de datos muy grandes
Percepción paralela	Exploración de sistemas de alta dimensionalidad, monitoreo de múltiples procesos, comparación de múltiples conjuntos de datos

Tabla 5.1: Ventajas que se obtiene en la utilizacion de un display acústico. (Adaptado de [Kramer 94]).

Resolución temporal de eventos	Datos que poseen información temporal (inclusive de milisegundos a milésimas de milisegundos)
Respuesta afectiva	Facilidad de aprendizaje, entretenimiento, posibilidad de expresar información afectiva o expresiva
Formación de *gestalts* auditivos	Discernir relaciones globales o tendencias en datos, percibir eventos importantes o estados en un stream de datos

Un interesante capacidad para aprovechar en DA es la buena resolución temporal que ofrece el sistema auditivo con respecto a la vista. El régimen de 25 cuadros por segundo de la televisión (sistema PAL) pone un límite a la resolución temporal, haciendo que eventos de menos de 40 mS no se puedan representar. Con cualquier hardware de sonido para computadora, se pueden representar y percibir eventos de hasta milésimas de milisegundos de diferencia.

Otro tema importante es la *correlación intermodal*. Esta característica se refiere a la convergencia sensitiva que ofrecen los sentidos en una experiencia determinada. Por ejemplo, la correlacion visual-auditiva puede verificar si dos objetos están en contacto entre sí. Las correlaciones intermodales o modalidades cruzadas implican redundnacia de información, pero esto provee un realismo incrementado en ambientes virtuales. En ambientes virtuales y en el caso de no videntes donde no hay feedback visual de la actividad desarrollada, las experiencias de interacción no son facilmente expresables.

Se denomina *sinestesia* a la sustitución de una modalidad sensorial por otra, y esta es una técnica utilizada para solucionar el problema anterior. Un ejemplo común en aplicaciones de VR y en el sistema HiperAudio es el siguiente: para proveer una sensación que reemplace el feedback táctil generado al tocar un objeto virtual, el sistema genera un sonido especial para dar el feedback indicando que el usuario ha tomado el objeto en cuestión.

La demanda de CPU y el costo de displays gráficos sugieren que soluciones efectivas a través de DA pueden ser viables, pues la carga computacional

demandada en compración con esas tecnologias es mucho menor. Por ejemplo un minuto de video sin compresión precisa 14 gigabytes para ser almacenado, mientras que un minuto de sonido stereo de alta calidad solo precisa 10 megabytes. La adición de sonido 3D, con efectos de difracción, difusión y reverberación ambiental, implica un incremento en el costo computacional que se traduce en un incremento en el valor final.

Dificultades existentes con los Displays Acústicos

Los DA también imponen dificultades y limitaciones. Estas complicaciones, en algunos casos resultan directamente de la manera en que nuestro sistema nervioso procesa el sonido, mientras que otros problemas pueden ser asociados con la tarea presentada o con el ambiente en el cual el DA es usado. La siguiente tabla resume los problemas en cuestión:

Desventaja	Explicación
Baja resolución de muchas variables auditivas	Difícil representación de datos cuantitativos
Limitada precisión espacial	Baja resolución espacial, representaciones volumétricas son representads pobremente
Falta de valores absolutos	La sonificación de datos es relativa, íconos auditivos indican estados generales del sistema, no valores precisos
Falta de ortogonalidad	Muchos parámetros auditivos no son perceptualmente independientes
Interferencia con comunicación oral	Trabajo grupal puede ser dificultoso
No hay limitación sólo a la linea de la vista	Oyentes cercanos pueden ser molestados, falta de privacidad
Limitaciones del usuario: equivalente aural de la ceguera al color	Como sucede con display visuales, algunas personas tienen menos resolución a variables tales como timbre, altura, ritmo, posición espacial, etc..
Ausencia de persistencia	Dificultosa revisión o comparación de dos

Tabla 5.2: Problemas que surgen en la utilizacion de un display acústico. (Adaptado de [Kramer 94]).

	regiones de datos
No hay impresión en papel	Distribución de resultados problemática

Un caso a tener en cuenta en nuestro sistema, el cual utiliza presentación a través de sonido 3D, es la baja resolución o precisión espacial del sistema auditivo. La resolución visual en el área fóveal de máxima resolución es tipicamente de un minuto de arco, llegando para tareas especiales hasta 2 segundos de arco. La capacidad del sistema auditivo para localizar una posición en el espacio es mucho mas grosera, con 1 grado de resolución en el frente del oyente y de 5 a 10 grados en los costados [Wenzel 92]. Cuando uno está representando datos indexados espacialmente, esta limitación es de particular importancia.

Sonificación

Sonificación es un nuevo campo, en el cual inclusive su nombre es aún motivo de debate. La siguiente definición es la que comunmente se maneja: sonificación es el mapping de relaciones representadas numéricamnete en algún dominio de estudio, a relaciones en el dominio acústico con el propósito de interpretar, entender o comunicar relaciones en el dominio estudiado [Scaletti 92].

La sonificación permite mapear diferentes variables de un cierto fenómeno a diferentes variables de una cierta señal acústica. Esta selección debe ser cuidadosa. Supongamos que hemos puesto un medidor de cantidad de autos que pasan por segundo en una autopista. Si esta medición se hace durante una semana, tendremos una serie de datos bastante extensa. Si con estos datos queremos construir una señal para ser escuchada y asi percibir la tasa de tráfico, la elección de como la señal va a ser generada es importante. Si cada valor registrado es mapeado a la amplitud de la señal, lo que obtendremos será ruido. En cambio si lo mapeamos a la amplitud de una señal constante, lo que escucharemos será una señal que aumenta propocionalmente en función de la tasa de autos que pasan. Si este proceso es comprimido, podríamos percibir una semana de datos (tal vez varias planillas de números) en sólo unos segundos. De esta manera con o sin feedback visual, tendríamos una idea si circulan mas autos en un dia de semana que en un fin de semana.

En general, la característica que queremos analizar en los datos, sugieren una técnica de síntesis particular.

Característica de los datos a sonificar	Técnica de síntesis sugerida
Oscilación entre diversos estados	Interpolación tímbrica
Ejes o grillas	Tonos fijos, resonadores
Comparación	Sumas, productos, diferencias, correlación
Texturas y tendencias	Síntesis granular, FM (modulación en frecuencia), *waveshaping*, histograma sónico
Detección de periodicidad	Datos como muestras de sonido
Objetos virtuales en un espacio de Realidad Virtual	Modelos físicos, sonidos digitalizados
Datos con cierta actitud	Sonidos de instrumentos musicales, escalas musicales, sonidos digitalizados

Tabla 5.3: Según la característica a analizar en un conjunto de datos, se pueden utilizar ciertas técnicas de síntesis para producir el mapping al dominio acústico .

La aplicación de esta modalidad de representar información tiene particular importancia en ciertos casos. Por ejemplo, dados dos conjuntos de datos prácticamente similares se pueden notar las diferencias usando técnicas de mapping de datos a frecuencia. En resúmen, cada valor de los datos obtenidos va modificando la frecuencia de una señal. Este proceso se realiza para los dos conjuntos y luego se suma punto a punto las dos señales obtenidas. Diferencias en los conjuntos de datos del orden del milésimo son detectados como un batido, diferencias del órden del diezmilésimo aparecen como un fenómeno de cancelación de fase, implicando un notable cambio de amplitud en la señal final. Una revisión de las técnicas aquí mencionadas se pueden encontrar en [Scaletti 92].

Las técnicas de sonificación podrían ser útiles para explorar rápidamente densidad de información. Hay que tener en cuenta que en un sistema sin pistas visuales uno no puede saber a priori la longitud de cierto chunk de información. Mapeando la longitud a la duración de una señal sería una alternativa para saber aproximadamente el largo de una cierta porción de información.

Diseño de un display virtual acústico 3D

Son bien conocidos los guidelines que permiten diseñar cierta GUI o diseñar una aplicación que corra sobre una GUI. Esta modalidad de interfaz gráfica, se ha tornado un exitoso elemento en la interacción hombre-máquina, permitiendo que múltiple aplicaciones usen simultáneamente un display visual y que el usuario interactúe con ellas aprovechando el ancho de banda que provee el sentido de la vista. Estas capacidades provienen realmente de la combinación de 2 características:

- El usuario puede controlar la organización espacial de múltiples objetos visuales (ventanas) por medio de cierta interfaz física y un manejador de ventanas
- El usuario puede cambiar su atención visual entre varios objetos representados

Con el incremento de la importancia de las aplicaciones computacionales y la posibilidad de brindar al no vidente una modalidad de interacción y representación de cierto sistema, se precisa tener las herramientas de diseño para presentar, manejar y organizar sonidos evitando una cacofonía confusa producto de múltiples fuentes de audio simultáneas.

En nuestra vida real, esta confusión no sucede normalmente pues nuestro sistema auditivo ordena los sonidos de acuerdo a su timbre y sus posiciones en el espacio. Es por esta razón que se torna importante la utilización de sonido 3D y de la manipulación de sus características.

Si bien desde [Ludwig 90] en donde se expresan las primeras ideas acerca de la extensión de un sistema de ventanas potenciado con sonido 3D para manipular más facilmente el ambiente y desde [Gaver 89] en donde se extiende el desktop de McIntosh para producir un feedback adecuado en función de la actividad del usuario, no han surgido resultados concisos en el diseño de displays acústicos, con más razón displays acústicos 3D.

Es por esta razón que no existen guidelines precisos para llevar a cabo esta tarea y por ello se debe recolectar el background existente que sumado a la experiencia de este trabajo plantean algunos puntos a tener en cuenta.

Guidelines a tener en cuenta

La metodología para diseñar un DAV (Display Acústico Virtual) no sólo involucra el dominio de la aplicación, sino que también incluye aspectos culturales, físicos y de implementación. La literatura es escasa con respecto al diseño de display acústicos sin pistas visuales. Por esta razón, parte del trabajo

original de este trabajo es resumir que modalides son interesantes de incluir en un DAV.

El diseño de un DAV cubre básicamente dos aspectos :

• el aspecto técnico de la manipulación de sonidos

• el aspecto estético y conceptual de la organización espacial de ellos

Estos dos puntos serán ejemplificados en el capítulo que trata la implementación del sistema.

Efectos

La radio, la música y el cine han utilizado efectos de sonido desde hace años para realzar un cierto efecto, dar importancia a un evento o realzar un sonido sobre otros. En particular la industria de la música utiliza comunmente procesamiento electrónico para enfatizar solos de instrumentos o vocalistas. Este procesamiento es útil para dar énfasis a señales individuales de audio.

En [Ludwig 90] se comentan algunos efectos interesantes y el resultado que producen:

Efecto	Resultado
Auto-animación	Incrementar variaciones de frecuencia en la señal original. Es el mismo efecto que originan pequeñas piedras en un rio turbulento, incrementando el patrón visual de la turbulencia
Distorsión	Producir un sonido deformado, como en el caso de las guitarra eléctrica de un músico de heavy rock. Este efecto reduce la inteligibilidad e incrementa la fatiga.
Filtrado	Filtrado con filtro pasabajos, pasaaltos o elimina banda, de tal manera de incrementar o disminuir cierta frecuencias, dando realce a cierta parte del espectro
Distanciamiento	Inclusión de ecos, reverberación y cambio en la amplitud para producir sensación de distancia.

Tabla 5.4: Para dar realce o acentuar alguna propiedad en general de una fuente sonora se puede aplicar diferentes tipos de procesamiento.

Resúmen

Normalmente el sonido es tratado como un complemento cosmético a una interfaz en un sistema de cómputo, pero cuando una señal visual no es factible de ser transferida al usuario, el sonido se torna el medio de transferencia más importante. El estudio de displays virtuales acústicos es una incipiente área dentro del campo de las interfaces. Existe un background existente proveniente de la sicoacústica y del legado que ofrece las interfaces visuales, pero la temporalidad del sonido y su expresividad ponen otro tinte al diseño. La manipulación de sonido de manera digital para incluir efectos es un interesante color en la paleta del diseñador, siendo el sonido 3D un importantísimo elemento a conjugar en el diseño de una interfaz enteramente acústica. No existe mucha bibiografía que hable de la utilización estética del sonido 3D, es por ello que el legado de este trabajo, es arrojar algunas ideas acerca del diseño de un display virtual acústico con sonido 3D.

Introducción a Hipermedia Acústica Tridimensional

Capítulo

6

Resúmen

Este capítulo se dedica a la descripción del sistema propuesto. En primera instancia se describe la idea del modelo conversacional como metáfora de representación de la información [Lumbreras 95a, 95b]. Se verá como se realiza el mapping del modelo hipermedial al modelo conversacional, y cómo el usuario interactúa con el ambiente. Para ello, el concepto de metáfora está permanentemente en juego.

Se revisarán los conceptos vigentes en una interfaz visual, los cuales serán mapeados o representados en nuestra interface acústica de manera adecuada.

Metáforas

Es bien sabido que las aplicaciones actuales, introducen un nivel creciente de complejidad. Cuanta más funcionalidad uno introduzca en un dispositivo, mayor cantidad de controles, procedimientos y restricciones son introducidos.

La explicación de un fenómeno o procedimiento a través de un ejemplo similar o de algo que pertenece previamente al framework cognitivo del usuario es denominado genéricamente metáfora.

Veamos un ejemplo concreto para ilustarlo: El caso de la videocasetera es notable, pues la cantidad de funciones que ella provee son generalmente desaprovechadas por el usuario final, debido a la complejidad que ofrece su operación. Por ejemplo: la grabación de un programa con antelación introduce

dificultades que deben ser generalmente solucionadas con el manual al lado del operador, y siempre dejando la duda si realmente grabará nuestro programa favorito esa noche. Una metáfora de linea del tiempo mostrada en pantalla, podría solucionar facilmente ello, mostrando que canal queremos grabar, y a que hora comienza y finaliza la grabación.

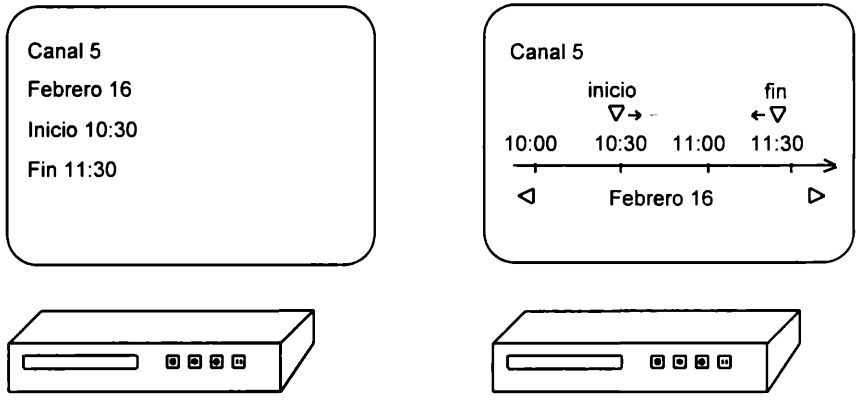


Fig. 6.1: Un ejemplo de lo útil que puede ser una metáfora: imagine la operación de programación para grabación automática de estas dos máquinas y evalúe cual intuitivamente es más fácil de operar.

Este ejemplo muestra un falla en el diseño de la interfaz del dispositvo, pues muchas videocaseteras que poseen OSC (On Screen Command) tienen la tecnología para soportar la interfaz propuesta y no lo hacen.

La idea de metáfora de línea del tiempo es una buena solución, pues expresa una idea común en el usuario medio: el tiempo transcurre hacia adelante y que cada instante es un punto en esa línea; un punto indica comienzo de la grabación y un punto a su derecha indica la finalización de la operación.

Es así que una metáfora es importante por dos razones:

• Puede ser usada para reducir la complejidad conceptual de una aplicación, haciendo parecer la interacción con esta a algo que ya probablemente existe en el framework cognitivo del usuario, y

• puede actuar como una poderosa ayuda en el diseño real de la aplicación computacional y en la representación de la interfaz de usuario del sistema

El diseño de una metáfora debería seguir el siguiente análisis en su diseño, y a modo de guideline tenemos los siguientes puntos a tener en cuenta [Erickson 90]:

• definición funcional: que hace el sistema y que puede hacer el usuario con él,

• identificación de los problemas del usuario: que aspectos de la funcionalidad son nuevos al usuario, que parece natural y que es diferente,

• generación de la metáfora: que hay implícito en la definción del problema?,

• evaluación de la metáfora, lo que implica verificar:capacidad de estructuración de contenido y aplicabilidad de la estructura (la metáfora es relevante al problema? , se maneja bien la expectativa del usuario de tal manera que el usuario puede esperar algo que nunca sucederá o estará diponible?)

• representabilidad: es fácil de representar e implementar la metáfora?

• grado de adecuación al usuario destino: el usuario final entenderá la metáfora?

• extensibilidad: que otras cosas puede soportar la metáfora para una futura funcionalidad extra?

Teniendo estos puntos en mente surge la pregunta: cuál es la metáfora adecuada para representar un sistema de información de naturaleza hipermedial con representación acústica, y para ser utilizado por un no vidente? Recordemos que la metáfora debe entonces satisfacer dos cosas:

• debe ser algo conocido por un usuario no vidente, y

• debe poder permitir mapear una hiperestructura en ella

Una conversación interactiva

Parecería difícil la elección, pero prestemos atención a la siguiente proposición: una conversación entre varias personas es un actividad que se presenta comunmente a un usuario, ya sea vidente o no. Además, una conversación generalmente trata con un tópico en particular, el cual es enfocado de diferentes

puntos de vista, dependiendo del locutor. Esta característica permite tipar la información. Además el flujo conversacional, sigue un caracter asociativo de ideas, teniendo subyacentemente una estructura "hiper" , pues las ideas se van enlazando entre sí en función del interés del grupo de discusión o del moderador a cargo. Cada idea se enlaza con la anterior de acuerdo a una cierta postura que puede asentir lo anterior, oponerse, ejemplificar, extender, etc. Tenemos así que cada link además puede ser tipado de acuerdo al tipo de conectividad e interacción entre los locutores.

Por lo anteriormente expuesto, vemos que una metáfora conversacional de varios locutores puede ser una elección interesante para nuestro sistema. Además veremos como cada punto del guideline anteriormente visto se va satisfaciendo a medida que la vayamos describiendo con más detalle.

Comparación de metáfora de página y conversacional

Los modelos hipermediales tradicionales plantean la idea de nodo como contenedor de información y *anchors* como objeto seleccionable para activar la funcionalidad de cierto link. Estos nodos, considerados como unidades posiblemente atómicas, contienen potencialmente varios anchors , representados éstos por medio de botones, hotwords o íconos. Usualmente en un sistema en el cual la información es presentada y seleccionada gráficamente, se presenta una página al usuario como unidad elemental de información. Además muchos de los sistemas de autoría (ToolBook, Hypercard, MacroMind Director, etc) y el standard HTML, hacen énfasis en la utilización de la metáfora de página con anchors como modelo de representación y diseño de la información.

Hay que notar que en el HiperAudio, el medio de presentación de informacion es el sonido, y este posee una característica temporal, eso quiere decir que una vez percibido este no deja trazo de sus existencia, es decir no posee una vida *a priori* indefinida en el tiempo, tal como lo puede hacer una página contenedora de información que es representada gráficamente.

Es por esta razón que el modelo subyacente que representa a una instancia específica del sistema HiperAudio cuenta con una sutil diferencia respecto a los modelos de implementación de hipermedia vistos corrientemente.

La diferencia surge de la siguiente situación: en una página con varios botones seleccionables, uno la percibe y puede seleccionar su próximo nodo destino. Con un medio temporal como es el sonido y la falta de pistas visuales que se prolonguen en el tiempo, es difícil pensar en como sería "clickear" sobre algo.

La conversación como modelo hipermedial

En nuestro sistema la selección del próximo link destino es posible cuando existe una ramificación en el flujo de la conversación. Sin entrar en detalles de la naturaleza de una conversación entre diferentes integrantes, podemos decir que en ella puede existir o no un moderador. Este locutor destacado, tiene la propiedad de regular el flujo de la información y asignar la palabra a los locutores. En nuestro sistema, existirá un asistente, que no oficiará como moderador, sino como un ayudante que permitirá realizar tareas de control.

Así, en una conversación sin moderador, cierto locutor puede intentar tomar el control de dos maneras:

• comenzando a hablar, interrumpiendo al locutor previo, o retomando el control luego de cierta pausa,

• realizando un breve comentario, y tomando posteriormente el control

Cada una de estas opciones presenta dos alternativas posibles y a la vez interesantes. La primera de ella se presenta cuando en la conversación existe un intercambio de ideas entre locutores, siguiendo una cierta línea de razonamiento o un argumento de discusión. Este camino que es generado en esta interacción entre locutores, podría ser percibida para un oyente al debate, como un camino predefinido o tour guiado en el tópico en cuestión.

La segunda alternativa para tomar el control funciona así: cuando cierto locutor finaliza una idea, un segundo locutor plantea un comentario que complementa, extendiende, ejemplifica o se opone al concepto anterior. En una conversación real, el control pasa al locutor que interrumpió si el grupo desea la información prometida. Ese asentimiento surge con acciones sutiles tales como asentimientos con la cabeza, giro de la cabeza, un comentario breve tal como un "...si...", etc. los cuales están ausentes en un sistema sin pistas visuales [Preece 92].

En nuestro sistema el usuario funciona como un locutor destacado: si en cierto tramo de la conversación existe un link a otro tópico de cierto tipo, el locutor encragado de presentar ese tipo de información, realiza un breve comentario. En ese instante, el usuario tien dos chances:

• descartar el comentario: como sucede en una conversación, si alguien interrumpe para decir algo irrelevante al contexto de interés del usuario, se desecha su comentario y la conversación sigue por cierto camino que sigue la idea desarrollada

• aceptar la interrupción: si el comentario genera espectativa positiva en el usuario, éste probablemente estará interesado en obtener el todo que fue prometido en el comentario. De alguna manera el usuario conoce el locutor y posee alguna manera de activarlo, ya sea por el teclado, joystick o guante

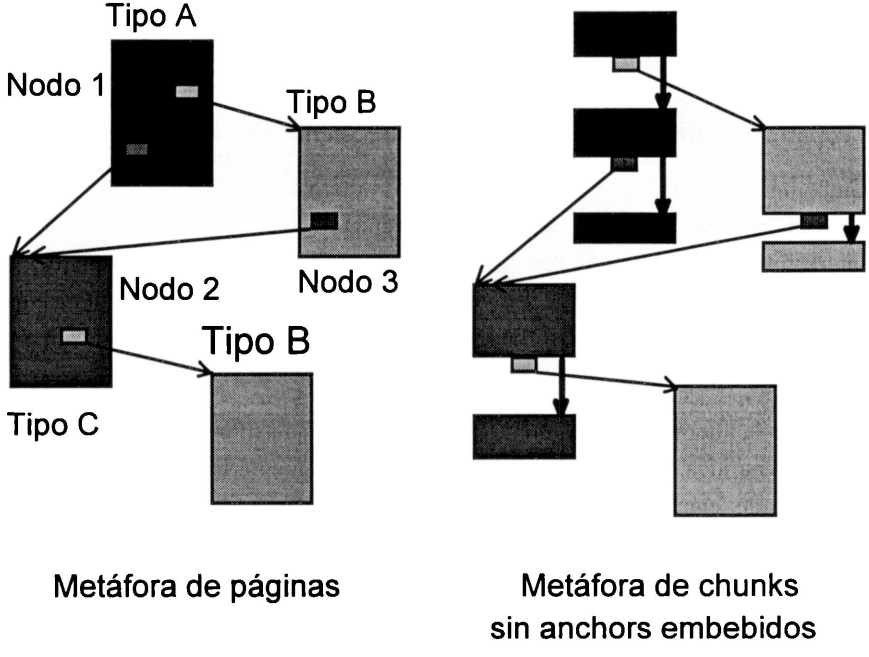


Fig. 6.2: Cada tono de gris representa un tipo de nodo. El nodo 1 usando la metáfora de páginas, se presenta al usuario en forma atómica, mostrando de alguna manera que éste posee dos links a los nodos 2 y 3. Esa misma estructura mapeada con el modelo de chunks sin anchors embebidos, reestructura al nodo 1 como tres nodos, con links a la finalización de los dos primeros y con links directos entre los segmentos generados. El link directo representa flujo predefinido en caso de desechar el link intermedio.

El modelo de interrupción como mecanismo de linking de información, no es sólo para pasar el control de un solo locutor a otro solo, sino que la idea es facilmente extendible. Esto quiere decir, que varios locutores pueden interrumpir simultáneamente. Luego de terminar el locutor original, cada locutor que posea un link a él, pronuncia un breve comentario. Así el usuario puede elegir el próximo locutor entre varios o dejar correr la conversación.

En este momento uno puede pensar que la idea de una página o segmento de información con anchors o en este caso con comentarios de links, podría ser representado como un segmento lineal de sonido, el cual posee segmentos

insertos (los comentarios) que representan los links. Pero este modelo puede ser simplificado si sólo admitimos segmentos o *chunks* atómicos sin links intermedios y que sólo poseen links a su finalización. En la figura 6.2 se ve como se representa la idea propuesta.

Modalidad de interacción en la conversación

Suponiendo que existe un cierto HiperAudio con las características propuestas deberíamos explicar cómo el usuario podría interactuar con el sistema. En HyperSpeech, por medio de reconocimiento de voz, el usuario puede pronunciar diversos comando para poder obtener el próximo nodo. Si el usuario pronuncia 'oposición' obtendría un nodo el cual se opone al nodo previamente escuchado.

Fig. 6.3: Sin una metáfora espacial, el usuario debe recordar comandos en un espacio cognitivo plano tal como sucede en HyperSpeech [Arons 91]. Si la voz de cada locutor es expresada monofónicamente prácticamente no hay idea de una representación espacial.

En el sistema propuesto, cada locutor es presentado a través de una voz particular, la cual fue digitalizada de diferentes personas reales. Pero además aparece un componente ausente en HyperSpeech: la posibilidad de dar representación espacial al ambiente por medio de sonido 3D. Esto quiere decir que la voz de cada locutor no va a ser presentada monofónicamente, sino que va a ser espacializada y presentada en una cierta posición del espacio. Más precisamente los locutores serán presentados en una forma semicircular en frente del usuario. De esta forma, cada locutor será el encargado de hablar de cierto tipo de información con su voz característica. Así se provee al usuario con dos pistas fuindamentales para manejar su espectativa: la voz del locutor y su posición en el espacio, pues la voz es reproducida por medio de sonido 3D.

Obteniendo ayuda en el ambiente: el asistente

En esta conversación el usario deseará ejecutar de alguna manera comandos y necesitará orientación espacial. Para ello existe una entidad destacada denominada el asistente. Este será el encargado de proporcionar y gestionar las tareas de control. El asistente no proporciona información pura tal como lo hacen los locutores, sino que permitirá interactuar con el sistema.

El asistente se presenta al usuario como una persona más, la cual da consejos específicos dependiendo el contexto de la informacíón, sirve como "pegamento" entre tramos a priori no conexos en la conversación, y conceptualmente presenta cada uno de las acciones posibles del usuario.

De esta manera el sistema se presenta consistente al usuario: los locutores presentan la información y el asistente gestiona el control. Asi el usuario siempre interactúa con personas, las cuales son personas virtuales.

El punto más intersante es la activación de las diferentes opciones que posee el usuario, y eso se lleva a cabo por medio de un concepto novedoso: íconos auditoriales 3D.

Controlando el sistema: íconos auditivos 3D

De la misma manera que existe íconos para representar gráficos abreviados,earcons e íconos auditivos hacen lo propio con el sonido. (ver capítulo 7).

Estos íconos auditivos representan una actividad específica, y por medio de una técnica de *drag & drop*, se puede activar su funcionalidad.

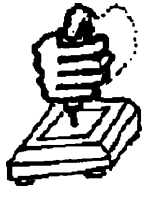


Fig. 6.4: La utilización de un joystick plantea una solución económica para manipular entidades en un ambiente virtual. Si uno imagina que está manipulando un cursor virtual en un ambiente tridimensional, podría seleccionar objetos en este ambiente.

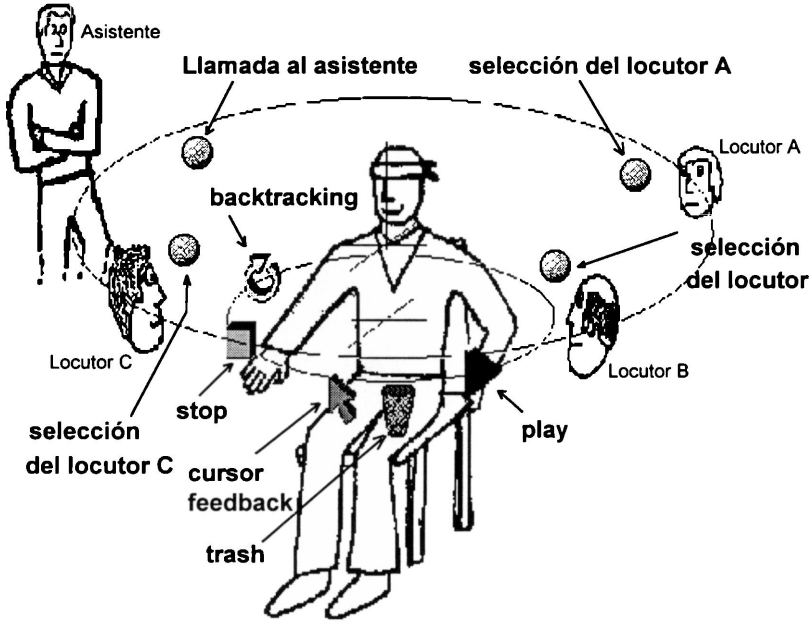


Fig. 6.5: Visión artística de una sesión con el sistema. Gráficamente se representan los íconos auditoriales, los cuales pueden ser tomados por medio de un guante especial. La modalidad cruzada kinéstesica-auditiva provista por el sonido 3D y el movimiento de la mano y el brazo reconocido por el guante, refuerza notablemente la sensación de espacialidad. La representación gráfica de los íconos es sólo ilustrativa, ellos son percibidos en modalidad auditiva. Un item interesante en particular, es el trash. Permite anular un ícono tomado erroneamente. En GUIs esto se hace clickeando en un área donde no hay nada.

Escalando información con una metáfora espacial: el edificio

Si bien la metáfora conversacional es una buena elección para la presentacion de información de naturaleza hipermedial, se presenta un problema con respecto al indexamiento y escalamiento de la información. Para resolver este problema tenemos dos alternativas:

• Extender la conversación: implica agregar niveles de presentación de información, mayor cantidad de locutores y tipos, etc. Esta opción complicaría las ventajas de la sencillez del modelo hasta ahora propuesto.

• Proponer múltiples conversaciones: si de alguna maner el usuario podría navegar alternativamente en diferentes conversaciones, en la que cada una trate un subtópico en general, estaríamos proveyendo una manera de escalar la información.

Esta última opción resulta interesante, pues si organizamos el espacio de información en diferentes conversaciones, las cuales están distribuidas en un cierta dimensión espacial, el usuario podría navegar espacialmente hasta encontrar la conversación de su interés. De aquí surgen varios problemas a resolver:

• cómo representar un espacio métrico de tal manera que este pueda ser navegado?

• cómo organizar las conversaciones en este espacio?

• qué metáfora es la más adecuada para esa organización?

• la metáfora elegida, cómo se complementa homogéneamente a la metáfora conversacional?

• que nivel de estructuración posee la metáfora propuesta?

Aquí se plantea el mismo problema que antes: cual es la metáfora adecuada?

No hay método directo para la elección pero pensemos en esta proposición: supongamos que cada conversación ocurre dentro de una habitación. Si estas habitaciones están dispuesta a lo largo de un pasillo, el usuario podría navegar acústicamente por este pasillo e introducirse en la habitación deseada. De esta manera, la disposición espacial de cada habitación podría oficiar a modo de índice. Si este piso de habitaciones es conectado entre sí a través de un ascensor que será comandado por el usuario, la organización de los pisos sería una estructura jerárquica para nuclear la información. El hall de entrada del edificio podría ser la tabla de contenidos de todo el HiperAudio y a su vez podría ser el contenedor del índice general. Es así que surge una metáfora que complementa a la conversación: el edificio. La conversación gestiona la interacción con la

información y el edificio organiza la estructura global. Sin embargo hay que tener en cuenta que los impedidos visuales sufren de una limitación en la representación de información espacial, exigiendo alta carga cognitiva para procesar estas estructuras [Hatwell 93]. Sin embargo es factible especular que este ambiente altamente interactivo promueva la construcción de robustos mapas de navegación en el usuario, tal como se puede extrapolar de [Canter 77].

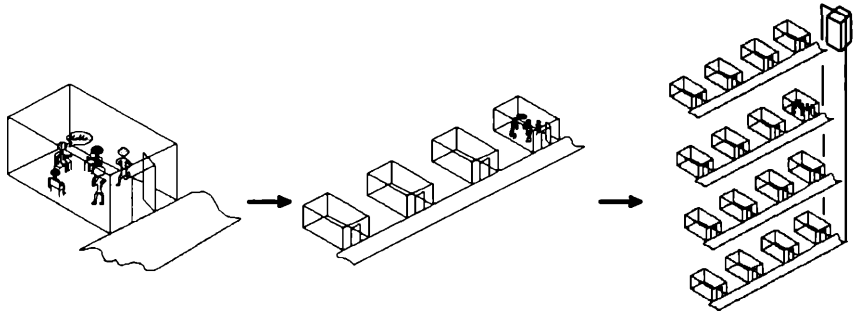


Fig. 6.6: Una conversación puede ser mantenida en una habitación. En cada una de ellas podría haber una conversación distinta. Estas habitaciones se pueden organizar de acuerdo a algún criterio en un pasillo, y estos se pueden organizar en pisos para dar una estructura jerárquica al espacio de información.

De esta elección surge un problema no explorado en la literatura de ambientes auditoriales que es el siguiente: cómo debe ser diseñado un ambiente espacial el cual va a ser navegado sin pistas visuales?

Resúmen

En HyperPhone [Muller 92] se sostiene que la navegacion se tiende a modelar espacialmente en casi todas las interfaces, y que la navegacion de voz es particularmente dificil de mapear al dominio espacial. Una hipermedia acústica tridimensional tal como la propuesta salva esa desventaja hasta ahora no resuelta claramente. Pero esta elección también involucra tópicos tal como selección adecuada de metáforas y acceso a información organizada espacialmente. Si bien el acceso conversacional ofrece el marco de interacción del usuario, tópicos tal como indexación y acceso a estructuras navegacionales implica nuevamente tratar con metáforas. Además, el usuario debe interactuar con el sistema homogéneamente y en forma continua, tratando de no incorporar modalidades de interacción artificiosas y aplicando en lo posible un criterio ecológico [Gibson 79]. Es decir, interactuar con el ambiente de la misma manera con la que uno interactúa en la realidad.

Este es el desafío que hay que batir y para ello HiperAudio se plantea como un prototipo que intenta validar las ideas aquí vertidas, y ser una plataforma para testear otras modalidades auditoriales para acceso a información.

HiperAudio

Capítulo

7

Resúmen

La implementación del sistema no solo incluye decisiones del diseño global de la arquitectura, sino que cubre tambien aspectos referidos a la interconexión de diversas piezas de hardware no convencional entre sí, tal como un HMD ,un guante para aplicaciones de realidad virtual y hardware con chip DSP para la generación de sonido 3D en tiempo real.

La implementación global del sistema es comentada, diferenciandose dos fases claramente distintas en el diseño y utilización:

- recolección, organización y edición de la información
- módulo de ejecución, el cual gestiona
    - sonido 3D,
    - interacción con el usuario,
    - funcionamiento de la máquina que maneja la información hipermedial

Diversas decisiones en el diseño de interfaz de usuario se comentan, así también como las diferentes alternativas testeadas para potenciales productos finales.

Descripción del sistema

El sistema se puede dividir en dos grandes partes: una es la que permite la autoría del HiperAudio y la otra es la que realmente lo ejecuta. Cada una de estas partes ofrece una modalidad de trabajo: la edición se realiza en modo off-line, obteniendo como producto final los segmentos de audio adecuadamente fragmentados y una descripción textual de la red de nodos y links ya editada. Esta descripción textual es obtenida por medio de un proceso de compilación realizado por el editor con un formato especial diseñado al efecto. De esta manera se separa claramente el producto de la edición y la fase de ejecución.

Con estos dos componenentes el modulo de ejecución ya tiene la información suficiente como para ejecutar la instancia de HiperAudio deseada. El proceso de obtención de información y procesamiento para la ejecución se puede observar en el siguiente diagrama funcional.

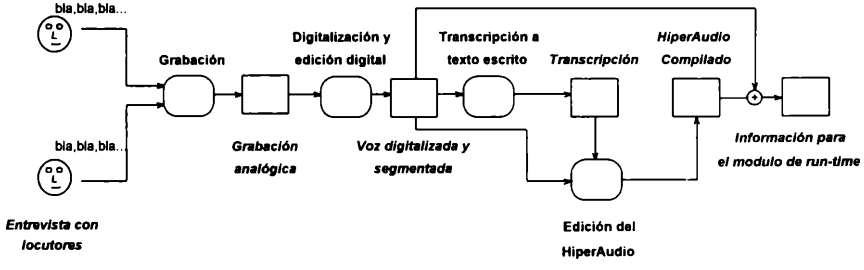


Fig. 7.1: El proceso de autoría involucra diversos pasos, siendo el mas complejo el de edición. Una vez que éste es finalizado, se genera una descripción textual de la red por medio de un proceso de compilación. Con esa descripción y los sonidos digitalizados, el modulo de ejecución recreará la instancia del HiperAudio deseada.

El modulo de ejecución es el encargado de leer al HiperAudio compilado y los segmentos de voz digitalizada de los locutores. Además gestiona el control de los dispositivos de hardware que permiten manipular al ambiente y presentar feedback al usuario. La figura 7.2 describe la arquitectura del módulo de ejecución.

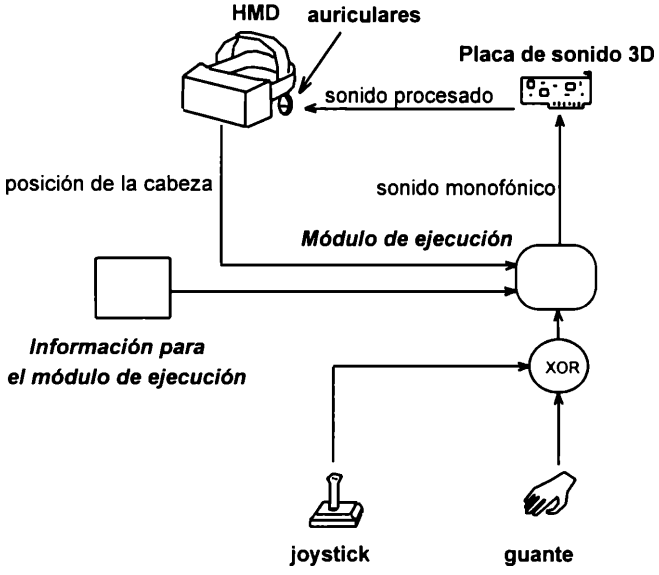


Fig. 7.2: El módulo de ejecución lee los archivos de sonido y la descripción de la red que conforma el HiperAudio. Además gestiona el hardware asociado a la interfaz de usuario. También controla la generación del sonido 3D.

A continuación se comentará en detalle como funciona cada una de las subpartes descriptas hasta aquí.

Obtención de la información

La información que presentará el sistema al usuario final está representada principalmente por voces digitalizadas de diversos locutores. Estos encarnan diferentes puntos de vista de un tema particular. Por ello, el primer paso en la generación de una instancia particular de HiperAudio es la obtención de la información. Esta captura se realiza grabando la voz de cada uno de los locutores que serán presentados finalmente en el HiperAudio. Existen dos chances para establecer el contenido a obtener:

- El locutor comenta el tema en una entrevista interactiva
- El locutor lee un *script* previamente confeccionado

La primera opción surge como las más fácil *a priori*, pues se obtiene información de forma casi directa, pero luego al recabar y organizar toda la información de todos los locutores surge que:

• Parte de la información no es relevante

• Cierta conocimiento que probablemente posee el locutor sería bueno incluírlo, pero este no está presente en la grabación

• Dependiendo del contexto acústico, puede haber tramos *sucios* por ruido ambiente

• Posiblemente ciertas afirmaciones podrían ser reestructuradas para resultar mas claras al usuario final

• Al estructurar la información, el autor descubre que links interesantes están ausentes debido a que falta alguna información que genere la conectividad argumental de la conversación. Hay que tener en cuenta que la conversación final se generará en funcion de las conversaciones aisladas, y que de antemano cada locutor no conoce el contexto de las otras grabaciones.

Por todo esto, es interesante generar un proceso de refinamiento el cual los locutores deberían satisfacer las máximas de Grice [Grice 75].

Tabla 7.3: Las maximas de Grice se refiere a la interacción entre partes para trabajos cooperativos. Igualmente son bien aplicables para la conversación virtual entre los locutores

Máximas de cantidad	Realice su contribución tan informativa como sea requerida No haga su contribución más informativa de lo que es requerida
Máximas de calidad	No diga lo que usted cree falso No diga algo para lo cual no tenga suficiente evidencia
Máximas de modo	Sea perspicaz Evite oscuridad en su expresión Evite ambigüedad Sea breve Sea ordenado
Maxima de relación	Sea relevante

Si los locutores pueden estar nuevamente disponibles, se grabarán nuevos segmentos para eliminar los problemas anteriores. Si bien la entrevista es un método directo de obtener la información, la generación de un script previo para el locutor muestra las siguientes ventajas:

• Organizar previamente el contenido de la información y simplificar los conceptos vertidos

• Acentuar en las aseveraciones el punto de vista del locutor, diferenciándose más claramente de los puntos de vista de los otros locutores

En función de la disponibilidad de los locutores, la generación del script es posible. La grabación puede llevarse a cabo directamente digitalizando la voz del locutor en la computadora, o grabandola con algún grabador portátil.

Detalles de la grabación y digitalización

En el prototipo presentado, la grabación se realizo con un grabador portátil tipo walkman AIWA HS-JS415, utilizando un micrófono omnidireccional condensador. El contenido de la cinta grabada, fue posteriormente digitalizado a 44Khz y 16 bits de cuantización. Esta elección fue realizada pues se trató de distorsionar lo menos posible la señal analógica, a pesar que la voz podría haber sido digitalizada a una menor frecuencia. La digitalización se intentó con tres placas de sonido para PC: Gravis Ultrasound, Gravis Ultrasound MAX y Turtle Beach Tahiti. La elección favoreció a la última, pues ofrecía menor relación señal ruido e introducía menor silbido, producto de oscilaciones propias del hardware de la PC. Es de notar que esta última placa cuesta alrededor del doble que las anteriores. El software utilizado para la digitalización fue el GoldWave V2.1 (shareware) , pues provee anulación de offset de la señal, manejo de fade-in y fade-out, facilidad de manejo de interfaz, etc. Se testeo también para este propósito: Cool Editor (shareware) y Wave SE (Turtle Beach). El programa Sound Edit (Macromedia), se probó en plataforma McIntosh. Este software ofrecía características interesantes tal como cambio de duración de una grabación sin cambio de altura del sonido. Lamentablemente la McIntosh Quadra 950 donde fue utilizado poseía una placa de sonido de 8 bits, haciendo ruidosa la digitalización.

La obtención de la información se realizó con una entrevista de aproximadamente 10 minutos. Unos dias antes, los locutores fueron notificados acerca de los tópicos a tratar en la entrevista.

Posteriormente, fueron citados nuevamente para grabar pequeñas frases que servirían de link a los conceptos vertidos anteriormente. Cada una de estas frases fue previamente confeccionada por el autor, y reflejaban cierta posición del locutor tal como: oposición, soporte, ejemplificación, extensión, etc. Los locutores no sabían como y en que contexto esas frases iban a ser utilizadas. Por ello algunas carecieron de la entonación adecuada.

Transcripción a texto

La información digitalizada, no es totalmente útil para la generación del HiperAudio. En especial el problema surge al intentar generar los links. Para poder realizarlos, hay que tener una acabada concepción de todos los *chunks* o fragmentos obtenidos en la grabación. Para ello el método utilizado es la transcripción a texto manual. De esta manera se puede cotejar las diferentes porciones de información y analizar en forma paralela diversos tramos de conversación buscando asociaciones y diferencias. La naturaleza transitoria del sonido y el pobre aporte de contendio de una representación gráfica de un sonido hacen que la transcripción sea el único método para obtener la semántica necesaria para generar los links. Si bien existen reconocedores de voz para limitar el trabajo de la tediosa transcripción, ese tipo de software es caro y no disponible generalmente en castellano. Además el estado del arte impone otras restricciones tal como dificultades de reconocimiento independiente del locutor y problemas para manejar un vocabulario arbitrario.

Edición

El editor de HiperAudio es una pieza fundamental en la autoría del sistema. El editor toma ya el sonido segmentado de cada locutor y la transcripción del texto. El sonido es utilizado para escuchar cada *chunk* de sonido durante la edición y la transcripción permite documentar la representación gráfica de la red.

Fundamentalmente el editor resuelve:

• Creación de tipos de nodos: uno para cada locutor

• Agrupamiento de nodos en unidades conceptuales llamadas grupos, los cuales permiten estructurar información con diferente grado de complejidad. Cada grupo posee una representación gráfica propia dibujada por el autor.

• Visualización y manipulación gráfica de la red de conceptos

• Ejecución del sonido asociado a cada nodo o link

• Compilación del HiperAudio para crear una descripción textual de la representación gráfica manipulada

El proceso de edición no sólo involucra la utilización directa de ciertas funcionalidades, sino que también involucra los siguientes tópicos:

• Estética de linking: no sólo la generación de links involucra asociar segmentos, sino que implica buscar una relación adecuada entre ellos, mantener una estética en el flujo conversacional, balancear el número de links que salen de cada nodo, generar caminos predefinidos a modo de *guided tour* de tal manera de conceptualizar una idea de varios locutores.

• Agrupación de nodos: cada entrevista con los locutores sigue un camino lineal. Pero ese camino va evolucionando, dejando trazas de diferentes tópicos que se van tratando a lo largo del tiempo. Si estas trazas son coincidentes entre varios locutores, los segmentos provenientes de ellas pueden ser agrupadas y dentro de ese grupo, linkeadas entre sí. La calidad del reportaje o del script favorecerá a la creación de grupos.

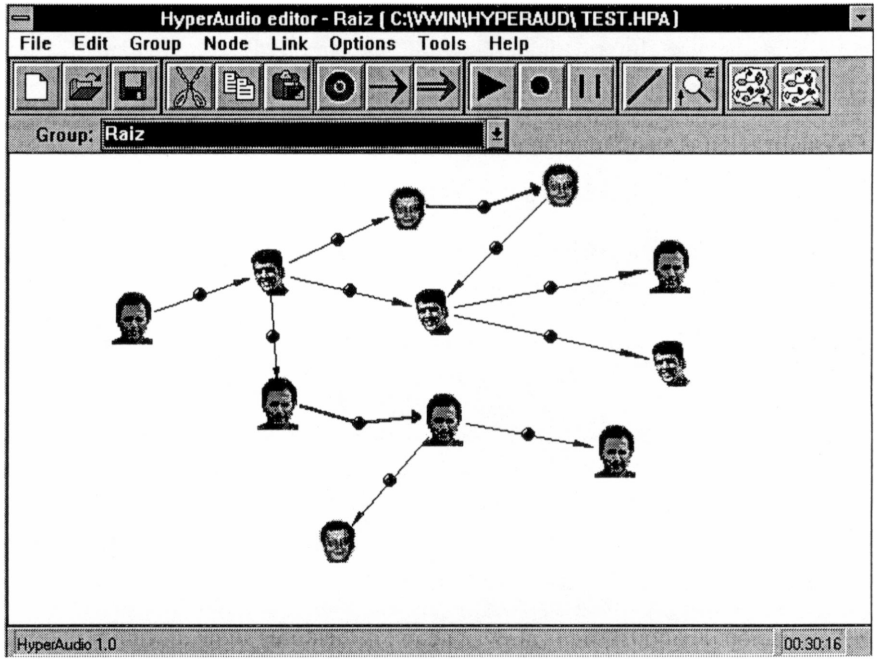


Fig. 7.4: En el siguiente snapshot se observa una instancia típica de edición de un HiperAudio. El editor permite crear tipos de nodos asociados a cada locutor y asignarle propiedades. El editor permite también digitalizar sonidos on-line si es necesario. Cada link se refleja con una linea con flecha. El comentario es representado por un círculo sobre el link. Las flechas con linea mas gruesa, indica un link especial: el link directo. Se muestra también la funcionalidad provista por la interfaz.

El editor en detalle

El editor es la herramienta principal en la autoría. Para indicar claramente su funcionalidad, veamos paso a paso como se genera y edita un HiperAudio. Veremos una secuencia típica de pasos. Esta secuencia no es estricta: por ejemplo una vez editado un HiperAudio es posible agregar un nuevo locutor inexistente originalmente, o por ejemplo se puede digitalizar on-line un sonido que previamente no existía. Los pasos serían:

Organización de la información

La voz de cada locutor ya fragmentada, se guarda en diferentes directorios: uno diferente por cada locutor. Esta tarea es de preparación para la edición y no es propia del editor.

Tipificación

Por cada locutor se crea un tipo de nodo especial, asignándole al él un bitmap para la representación icónica, un nombre, un directorio de sonidos y una posición deseada del locutor en el espacio. Esta información posteriormente será

utilizada por el módulo de ejecución para la presentación de sonido 3D. Esta posición no tiene nada que ver con la representación gráfica espacial del nodo en la pantalla de edición.

Documentación de la transcripción

Cada segmento de voz posee un nombre y una extensión WAV. Acompaña a cada segmento, un archivo de texto con el mismo nombre que el sonido pero con extensión .TXT , el cual internamente posee un título y un contenido con un formato propio. En este archivo un caracter especial separa el título del contenido. Su creación se hace con un editor de texto común. Esta información es posible visualizarla posteriormente en pantalla para documentar la red, indicando textualmente el contenido del nodo y su título.

Creación de nodos y links

Al crear un nodo este aparece en la pantalla en la posición deseada, la cual puede ser modificada por medio de manipulación directa. Con doble click sobre el nodo, una caja de diálogo permite asignarle un tipo y un archivo de sonido determinado. Seleccionando el nodo orígen con el boton izquierdo del mouse y el nodo destino con el botón derecho, se puede crear un link entre nodos. Estos links pueden ser links simples o directos. Esta característica será comentada posteriormente.

Revisación textual

Mostrar y ocultar el título del nodo, el contenido de texto o el nombre del sonido asociado permite editar la red, observando como los tópicos se van enlazando entre sí.

Ejecución y grabación de sonidos

Se puede ejecutar algún archivo de sonido para verificar el contenido. Esto se hace seleccionando el nodo y activando el ícono de ejecución. Además se puede grabar un sonido en en momento de la edición para completar la red. Esta funcionalidad está directamente soportada por el editor, seleccionando frecuencia de sampling y cantidad de bits de cuantización.

Reorganización gráfica

Moviendo los nodos en la pantalla con técnica de drag & drop se proporciona un método fácil para la organización gráfica de la red. El redibujo de los links se genera en forma automática.

Agrupamiento

Si es necesario y la red supera una cantidad apreciable de nodos, se pueden agrupar estos en subgrupos. De esta manera se maneja la complejidad de la red y se organiza conceptualemente en *clusters* o grupos a la conversación. Estos grupos se organizan en forma jerárquica. Cada grupo posee un nombre y puede llevar un ícono característico.

Salvado

Para no perder el trabajo, al terminar se puede grabar el HiperAudio. Esto se realiza bajo un formato especial hecho a tal efecto. En la grabación, queda embebido en el archivo la información de tipos, bitmaps de locutores y grupos, etc. además de la configuración gráfica del HiperAudio. De esta manera, cada archivo posee toda la información para ser posteriormente recreado en las mismas condiciones que se encontraba al ser grabado sin precisar otra cosa que ese único archivo. Una importante característica es que el *layout* gráfico de los nodos y links también se conserva en la grabación.

Compilación

Una vez finalizada la edición, se elige el nodo de inicio del HiperAudio, y se procede a compilar, generando una descripción textual de la red. Esta descripción se realiza con un lenguaje especial.

Compilación de un HiperAudio

En el capítulo 6 se comentó la diferencia en el diseño subyacente de una aplicación hipermedial basada en la metáfora de páginas con respecto a una metáfora de chunks atómicos. Este detalle probablemente no aparece en la etapa de diseño, pero en la fase de codificación debe existir una manera de especificar claramente el conexionado de cada una de las piezas de información.

De la misma manera que el estándard HTML permite describir cierta información para ser accedida de manera hipertextual, separando presentación del contenido, el editor de HiperAudio genera algo conceptualmente similar. Luego de compilar un HiperAudio se obtiene un archivo de texto, el cual responde a la siguiente grámatica BNF:

<HiperAudio> ::= [chunk_type]$^{+}$

<chunk_type> ::=<chunk> |

<chunk_with_buttons> |

<chunk_plus_seq> |

<chunk_plus_buttons_plus_seq> |

<button>

<chunk> ::=
CHUNK <chunk_id> <sound_file> POSITION <position_id> END

<chunk_with_buttons> ::= CHUNK+BUTTON <chunk_id> <sound_file> POSITION <position_id> BUTTONS [button_id]$^{+}$ END

<chunk_plus_seq> ::= CHUNK+SEQ <chunk_id> <sound_file> POSITION <position_id> CONTINUE AT <chunk_id> END

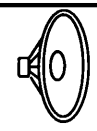
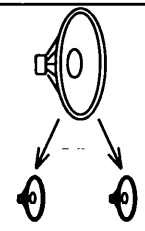
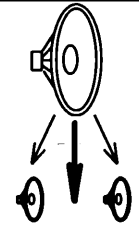
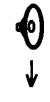
<chunk_plus_buttons_plus_seq> ::= CHUNK+BUTTONS+SEQ <chunk_id> <sound_file> POSITION <position_id> BUTTONS [button_id]$^{+}$ CONTINUE AT <chunk_id> END

<button> ::= BUTTON <button_id> <sound_file> POSITION <position_id> CHUNK DEST <chunk_id> END

<chunk_id>	::= integer	*{es un número único identificador de un chunk de sonido}*
<sound_file>	::= string	*{es el archivo de sonido con el path completo }*
<position_id>	::= integer	*{ es una posición discreta y abstracta de la posición del espacio en donde debe ser puesto el sonido }*
<button_id>	::= integer	*{ es un número único identificador del botón que representa un link }*

Con estos tipos simples de chunk es posible modelizar cualquier HiperAudio. La semántica de cada chunk aparece en la siguiente tabla:

Fig. 7.5: Una conversación interactiva que involucre al usuario como moderador, seleccionando links y siguiendo paths predifinidos o guided tours es fácilmente modelizable con estos constructores. Practicamente cualquier HiperAudio es especificable con estas entidades. Una característica ausente en hipermedia tradicional es el link directo, activado automáticamente luego de un time-out.

Representación conceptual	Funcionalidad
	**Chunk** Es un sonido simple que es ejecutado en una posición determinada del espacio
	**Chunk plus buttons** Idem anterior, pero al finalizar su ejecución, se ejecuta cada uno de los sonidos asociados a cada botón. Si luego de escuchar un botón el usuario lo activa, se navega hacia el chunk destino indicado por el botón
	**Chunk plus sequence** Idem al chunk , pero al finalizar su ejecución se accede directamente al chunk destino indicado.
	**Chunk plus buttons plus sequence** Idem al chunk plus buttons, pero si luego de un time-out el usuario no seleccionó ningún boton, se accede directamente al chunk apuntado por el link directo.
	**Button** Ejecuta un sonido asociado. Si el botón es activado, se accede al nodo destino apuntado por el botón.

En este modelo no se permite *a priori* links embebidos en un chunk, tal como aparece normalmente en una metáfora de páginas. Para salvar esta característica, el chunk original debe ser particionado, tal como muestra la figura 6.2.

Otro detalle interesante, ausente en aplicaciones hipermediales que no poseen noción de temporalidad, es el denominado link directo. Este link aparece en el chunk_plus_buttons_plus_seq. Luego que el usuario escucha el chunk y los sonidos de cada uno de los botones, comienza a correr un timer. Al expirar este y no habiendo selección del usuario, se activa el chunk destino. En el caso del chunk_plus_seq, el timer se inicializa con cero, indicando acceso inmediato al chunk destino. Tópicos referidos a temporalidad en documentos hipermediales y sincronización aparece en [Buchanan 92].

Otro tema a tener en cuenta es el agrupamiento. Si bien el editor permite manejar grupos, organizados jerárquicamente, esta representación es solo conceptual. En el proceso de compilación, todos los grupos son desanidados. De esta manera todos los chunks pertenecen a un solo grupo plano. Sin embargo el autor en edición maneja grupos sin enterarse de este detalle.

Ejecución del HiperAudio

El producto final de la edición, es un archivo de descripción (HiperAudio compilado) y el conjunto de archivos de sonidos. La ejecución del HiperAudio involucra la gestión del hardware y manejo de interfaz de usuario. La versión con HMD y hardware de sonido 3D de tiempo real es la propuesta de máxima. Un producto final de más bajo costo implicaría sacrificar ciertas posibilidades de interacción pero con una amplia base de plataformas instaladas. Si uno procesa de manera off-line todo el sonido, para generar sonido 3D, y cada uno de los íconos auditoriales son procesados para todas las posiciones posibles, queda un producto de accesible plataforma. Lo que se quiere indicar es que con una placa de sonido stereo corriendo bajo MS-Windows, auriculares y un joystick se puede ejecutar un HiperAudio. El problema de la especificidad de cada placa de audio es manejada por el driver de Windows , eliminando problemas propios del hardware. Como medio de soporte de la información el medio más barato seria un CD, el cual podría almacenar hasta 72 minutos de HiperAudio con sonido 3D. El punto a tener en cuenta ahora sería: Cómo se puede generar sonido 3D de manera off-line?

Procesamiento off-line de sonido 3D

Revisando los conceptos de los capítulos 2 y 3, la idea fundamental para obtener sonido 3D es la siguiente: convolucionar la fuente de sonido monofónica con el filtro FIR de cada oído, adecuado a la posición del espacio deseada. Sin complicaciones ese filtro FIR puede tener incluida también la informacion de ITD e IID, junto con la HRTF. La cuestión es la siguiente: Como obtener esos filtros?

La primera alternativa fue tratar de medirlos de modo casero siguiendo lo expuesto en la figura 2.15. La primer prueba consistió en armar un pequeño preamplificador que usara dos micrófono electret pequeños. Estos micrófonos serían puestos en el canal auditivo de alguna persona y así tratar de grabar un sonido para ver que ocurre. La primer prueba consistió en grabar sonidos del ambiente con esa modalidad y posteriormente intentar su reproducción con auriculares. Al reproducir se percibe un interesante efecto tridimensional, tal como lo anticipaba [Begault 94]. Eso es debido a que la grabación ya captura todos los parámetros (ITD, IID y HRTF). A pesar de que el experimento es casero, funciona. Ahora el tema sería obtener las HRTF. Para ello y tratando de ver "la firma" frecuencial que impone el oído, intentaría grabar un pulso con este método para obtener la función de transferencia del óido. De vuelta al método casero, la generación del pulso provino de la explosión de un globo de látex inflado, tal como lo describe [Begault 94]. Las no linealidades del sistema de grabación, la impureza del pulso obtenido que involucra inclusión de frecuencias no deseadas, la pobre acústica del lugar de grabación y el ruido de cuantización y de línea, hacia que la HRTF obtenida sea bastante pobre. Además en la función de transferencia se incluían los ecos de la sala de grabación, pues no era una cámara anecoica. En definitiva el intento era válido, pero los medios eran técnicamente menos que decentes para una calidad aceptable. La chance que quedaba era obtener HRTF's profesionalmente obtenidas.

Sets de HRTF's obtenidas

Luego de una revisión bibliográfica el único proveedor posible parecía ser el profesor Fred Wightman del Waisman Center, Universidad de Wisconsin, Madison, USA. [Wightman 89]. El profesor Wightman ofreció un completo set de HRTF's . Este set venía como un stream de mas de 2Mb de números y una indicación de como interpretarlos. Luego de particionar la información laboriosamente, se organizaron 144 pares de filtros, particionando el espacio circundante en 144 posiciones en la esfera virtual que rodea al oyente, con 24 posibles azimuth (cada 15°) y 6 posibles elevaciones (desde -36° a 54°). Los filtros estaban sampleados a 50Khz 16 bits. Para poder ser utilizados en calidad CD (44Khz) y menor (22Khz) hubo que resamplearlos siguiendo la técnica descripta en [Wheeler 93]. Para ello también se escribió un software adecuado.

La condición era que el set del Dr. Wightman debia ser usado para fines academicos y no con provecho comercial, firmando previamente un convenio a tal efecto. Posteriormente y confrontando datos, la empresa más importante de sonido 3D, Crystal River Engineering, fabricante de hardaware para sonido 3D tal usa uno de estos sets en el Convolvotron, Beachtron y Alphatron.

Otro set obtenido es provisto libremente por el el MIT via ftp anonimo en el host sound.media.mit.edu directorio /pub/Data/KEMAR. Este conjunto no es medido como en el caso anterior de personas reales, sino que es obtenido de un maniquí KEMAR de características antropomórficas fabricado por Knowles Electronics Inc. Itasca, Illinois, USA. Este set era más completo pues barría toda la esfera virtual del potencial oyente.

Con estos datos en manos, la tarea es realizar un módulo de convolución y testear.

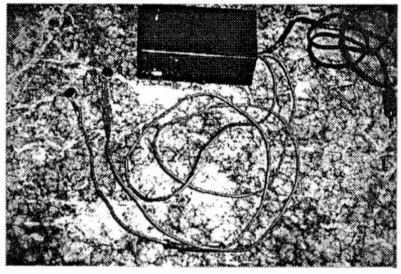


Fig. 7.6: Fotografía del preamplificador (caja negra) y los dos micrófonos electret para ser introducidos en los canales auditivos de una persona. Si bien el experimento es similar al que realizan profesionales, problemas técnicos de acústica y ruido hacían que las HRTF's obtenidas no sean usables.

Generador de sonido 3D off-line

Para efectuar el procesamiento off-line se escribió un programa especial. El sonido a procesar se divide en frames de medio segundo, y por medio de un script en donde se especifica el sonido que uno quiere procesar junto con la posición deseada de cada frame en el espacio. Con esta información el programa genera el sonido 3D.

La ventaja de esta aproximación es la calidad del sonido obtenido. Por problemas de potencia del hardware, el procesamiento on-line exige filtros acortados, es decir de menor cantidad de taps. Esto va en desmedro del filtrado final. En cambio en la versión off-line se puede utiulizar filtros de largo completo. El programa escrito en Turbo Pascal V6.0 se fue optimizando hasta obtener un segundo de sonido 3D por cada minuto de procesamiento, para filtros de 256 taps.

Resultados

Probando los dos conjuntos, el del Waisman Center producía resultados más realísticos. Esto puede ser pues probablemente la modelización de la cabeza por parte del maniquí no era tan aproximada como la de la cabeza real de un ser humano. Probablemente detalles en la grabación también influyeron en el resultado final.

Los sonidos obtenidos estaban ya compensados en los filtros para aurículares Senheinser HD540. El Dr Wightman sugiere que a falta de estos se utilice uno con la respuesta mas plana posible. Se testearon empíricamente modelos tipo TWIN TURBO Sony, AIWA (intra canal auditivo) y comunes (SONY CD30). En general el ranking de mejor a peor obtenido fue: SONY CD30, AIWA intra canal auditivo y SONY TWIN TURBO.

Hay que tener en cuenta que la función de transferencia del auricular, la diferencia de la HRTF del oyente final y la HRTF usada y las no linealidades del sistema de reproducción afectan el resultado final. Para verificar el efecto generado luego del procesamiento, se testea la misma fuente monofónica, generando una señal stereo alterando el volúmen de cada canal. Este procesamiento arroja un sonido sin externalización y con localización sólo en el eje interaural de la cabeza. El sonido 3D provee mayor realismo, externalización y posibilidad de localización con un previo entrenamiento.

Este software de procesamiento permite entonces generar un HiperAudio en las condiciones de bajo costo descriptas anteriormente. A continuación veremos como funciona la versión de procesamiento de tiempo real.

Ejecución de un HiperAudio con procesamiento en tiempo real

La alternativa más interesante es la generación de un ambiente de alta interactividad, con manipulación de entidades virtuales acústicas espaciales, manteniendo *tracking* o seguimeinto de la posición de la cabeza del usuario. Esta última alternativa permite que cuando el usuario no reconoce claramente la posicion del espacio de una fuente de sonido, moviendo la cabeza es capaz de desambiguar la posición. Además la posibilidad de usar un guante de aplicaciones de realidad virtual, refuerza la sensación espacial, pues la modalidad cruzada kinestésica-auditiva recrea una vívida sensación espacial.

El hardware

Uno de los aspectos más interesantes es la interacción con hardware no convencional. Este podría ser el punto de vista del usuario, pero la programación de un sistema que los involucre no es una tarea trivial. El siguiente gráfico muestra como se conectan estos dispositivos a una PC

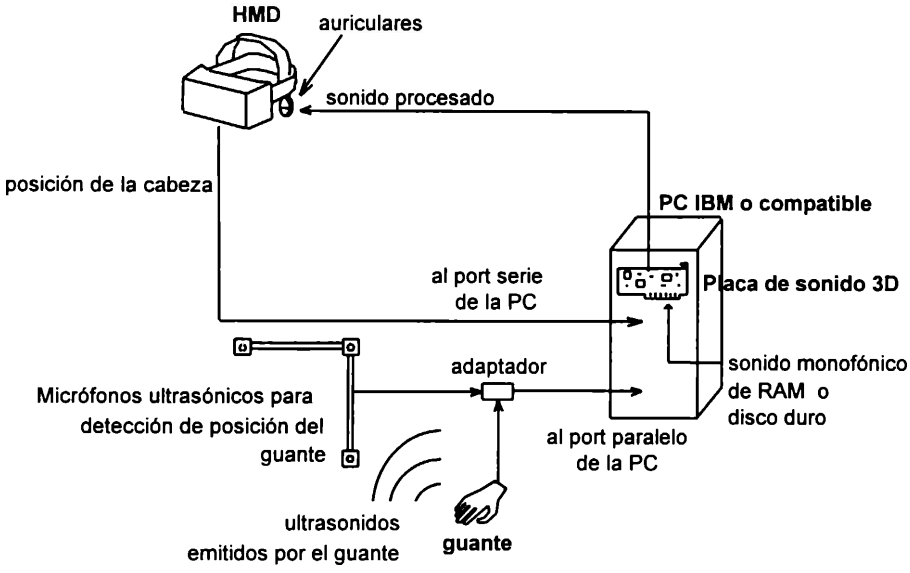


Fig. 7.7: Para poder usar todo el hardware la PC muestra una configuración fuera de lo común. El guante reporta su posición y status de los dedos utilizando las líneas de control de entrada del port paralelo. La posición de la cabeza se envía a través del puerto serial. La placa de sonido 3D recibe el sonido monofónico y lo procesa. Tanto hardaware exótico exige una cuidadoso diseño del software para evitar conflictos.

Veremos cómo funciona cada pieza de hardware y que características posee.

Hardware funcionando

El guante

El guante es un Nintendo Power Glove. Originalmente fue fabricado como reemplazo en el control de los juegos Nintendo. El guante empleado para esa función no fue nunca un éxito comercial. Su fabricación había sido discontinuada, pero ahora es retomada por una empresa norteamericana que compró la licencia y anunció su relanzamiento para febrero de 1996. Su costo es bajo (~ U$S 150).

El guante posee sensores de cintas resistivas en los dedos (excepto el meñique) discretizando el valor de cierre en 0,1,2 o 3. Por medio de dos emisores ultrasónicos puestos en la parte superior del guante y con un marco de 3 micrófonos ultrasónicos puestos al frente del usuario, se permite detectar la posición del guante en el espacio con aproximadamente 5 mm de resolución en un cubo de 2m x 2m x 2m. Para ello se utiliza una técnica denominada "tiempo de vuelo" [Kowalski 93]. En función del retardo con que arriba la señal emitida por el guante a los micrófonos se puede determinar su posición. Además se puede detectar el giro de muñeca con 30 grados de discretización. El guante no provee feedback táctil.

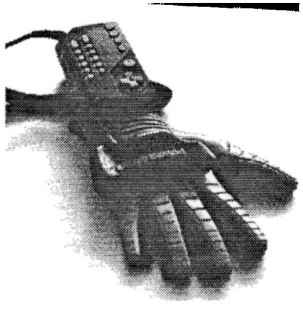


Fig. 7.8: Este es el aspecto del guante PowerGlove. Esta construido de plástico flexible en su parte superior y el guante propiamente dicho es de lycra. Además posee un teclado numérico que puede ser utilizado simultáneamente y leído desde la PC.

El guante viene provisto con un conector para los juegos Nintendo. Por esa razón un adaptador que posee latchs, microcontrolador, etc. adapta la señal para ser enviada a la PC a traves del port paralelo. Este port posee varias lineas de entrada de datos usadas normalmente por la impresora (busy, out of paper, etc.) Utilizando ellas, la PC puede recibir la información del guante.

Fig. 7.9: Este es el CyberMaxx visto de frente. En la unidad rectangular que finalmente queda contra la frente del usuario están los LCD y los magnetómetros. Unas correderas plásticas permiten ajustar el HMD y los auriculares a la posición deseada

El HMD (Head Mounted Display)

Un HMD es un dispositivo que integra un display, una óptica, tracking de la posición de la cabeza y auriculares en una sola unidad. Este "casco" es utilizado en aplicaciones de realidad virtual para mostrar imágenes stereo y sonido simultáneo. El HMD utilizado es fabricado por la empresa VictorMaxx y el nombre del producto es CyberMaxx. Posee dos LCD color de 120.000 pixels cada uno y tracking de la posición sin fuente emisora. Eso quiere decir que no necesita una referencia fija. Para ello utiliza magnetómetros que detectan el campo magnético terresetre!. En nuestro caso, no se utiliza las pantallas de LCD, pues el HMD sólo se lo utiliza para el tracking de la posición de la cabeza y los auriculares para la reproducción de sonidos. Lamentablemente, éstos últimos no son de buena calidad. En caso de recibir una imágen de la PC, por medio de un residente especial, se cambia el timing de la placa VGA de la PC para emitir señal con timing NTSC. De esa manera, se pueden observar imágenes en el casco usando para la generación una simple placa VGA. También el casco acepta video compuesto NTSC como señal.

El casco reporta la posición de la cabeza con tres grados de libertad (giro, elevación e inclinación), no reportando posición absoluta en el espacio. La información llega a la PC por el puerto serial, a 19200 baudios.

La placa de sonido 3D

Para este fin se probaron dos versiones de hardware. El primero fue una placa Gravis Ultrasound, de orígen canadiense. Esta placa es una placa de bajo costo ISA para PC y posee ciertas capacidades para la generación de sonido 3D en tiempo real. Utiliza una tecnología proveniente de la empresa Focal Point [Gehring 90]. Básicamente la idea es preprocesar el sonido de manera off-line para seis posiciones del espacio (frente, atrás, derecha, izquierda, arriba y abajo) y luego en tiempo de corrida interpolar proporcionalmente con las tres posiciones más cercanas a la posición finalmente deseada. Si bien este hardware

funciona, exige un crecimiento en el tamaño de los archivos de 6 veces y provee una calidad media de reproducción.

En la versión definitiva se utiliza la placa menor de la línea que fabrica Crystal River Engineering. En realidad esta placa es una Tahiti Turtle Beach. Esta placa posee RAM incorporada y un chip DSP Motorola 56001. Crystal River Engineering es pionera en sonido 3D y vende el software para programar el DSP junto con varios set de HRTF's seleccionados. Esta placa tiene la capacidad de espacializar en tiempo real sonidos de 16 bits y 44 Khz (2 fuentes) o 22Khz (4 fuentes simultáneas). Además se puede incluír efecto Doppler. Se pueden conectar hasta 6 placas en una misma PC para obtener mayor cantidad de fuentes sonoras simultáneas. [Alpha 95]. Crystal River Engineering vende otro hardaware tal como un server de sonido 3D "high end". Este server incorpora placas de hardaware denominado Convolvotron, el cual es capaz de simular reflexiones de la señal de audio en las paredes del ambiente con selección de diferentes materiales en las paredes, los cuales otorgan diferentes grados de absorción. Este server denominado Acoustetron, recibe información acerca de la posición deseada de las fuentes de sonidos, y devuelve a traves de una linea de audio el rendering adecuado de esa fuente. Técnicas acerca de los severs de audio 3D son tratadas en [Burgess 92]

Programando el hardware

Después de una ardua tarea de recopilación de toolkits, interfaces de programación y documentación, el software que controla el hardware debió ser escrito en C++. El principal motivo no fue una decisión de gusto, sino que la libreria de la placa de sonido 3D no se podía adaptar a ningún otro lenguaje, tal como Pascal.

Una de las características deseables sería la concurrencia entre la ejecución de un nodo y la interacción con la interfaz. Esto es una situación que potencialmente puede suceder y es factible que sea común. Por esta razón se debe realizar en forma concurrente y en tiempo real la mezcla de sonidos. Esta tarea se lleva a cabo a nivel de hardware por medio del llamado a funciones adecuadas de la librería provista por la placa Alphatron. Otro item a tener en cuenta es la dealocación dinámica de sonidos. Debido a que el DOS no es un sistema operativo multitarea, se debe realizar un esquema de polling explícito para saber en que momento un sonido a terminado y así gestionar su dealocación de memoria. Esta situación surge pues en el mismo momento que el sistema gestiona el manejo de la interfaz y activación sonora de íconos, puede

ocurrir que un chunk de sonido finalize. Para la ejecución concurrente de la maquina hipermedial y el gestor de interfaz, existe un loop que obtiene la actividad del usuario, actualiza la interfaz y gestiona la actualización del estado de la máquina hipermedial.

La arquitectura del software de la máquina hipermedial es *data-driven*, con todo el conocimiento embebido en los links y nodos, permitiendo que el software que recorre la red sea simple y conciso.

A continuación se observa un diagrama estructural del módulo de ejecución.

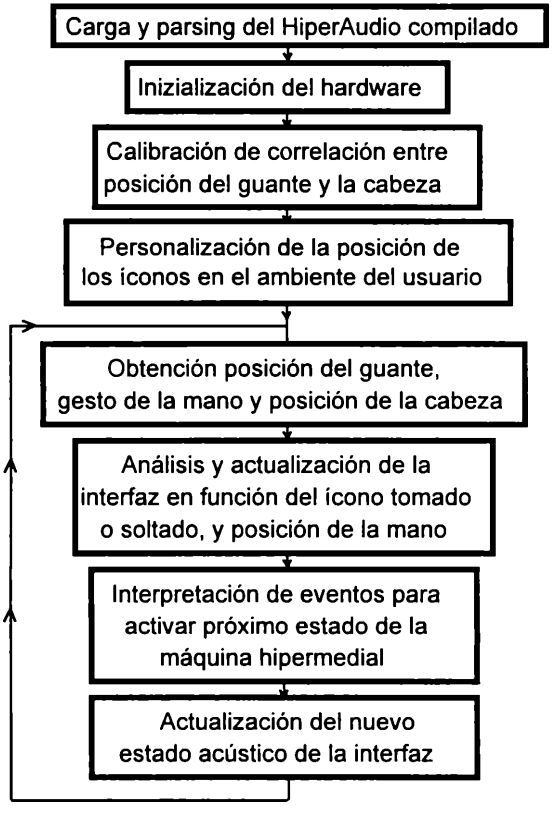


Fig. 7.10: Básicamente el módulo de ejecución lee el script generado por el editor, calibra la interfaz y permite la interacción con la máquina hipermedial, esta presentada al usuario a traves de la interacción con los locutores y los íconos auditivos. En la inicialización el usuario dispone los íconos auditivos en la posición mas cómoda para él. Además, se calibra la distancia entre la mano y la cabeza para un adecuada presentación acústica.

Una de las características interesantes en la calibración es la detección de la distancia entre el oído (centro de la cabeza) y la mano. Esto ocurre pues el guante en la inicialización toma como orígen de coordenadas cierto punto del espacio, el cual no es el centro de la cabeza. Este punto normalmente esta frente al pecho o donde haya puesto la mano el usuario en la calibración. Ahora si

nosotros ponemos un sonido "pegado"al guante y acercamos este al centro de la cabeza del oyente, el sonido no será adecuadamente presentado. Esto ocurre pues el guante cerca de la cabeza no devuelve (0,0,0) sino devolverá (offX, offY,offZ). Estos desplazamientos se refienen a la distancia que hay desde el centro de calibrado del guante hasta el centro de la cabeza. Así si sumamos esos offset a la posición actual del guante, obtendremos los sonidos 3D adecuadamente presentados. Para ello el usuario pone su mano frente a la linea de la vista, calibrando dos coordenadas (X y Z) y luego pone su mano al costado del cuerpo en el plano vertical de los dos oídos. Asi se calibra la "profundidad" o coordenada en Y. Esta modalidad de calibración es conocida como *ratcheting* [Hinckley 94].

Diseño de la interfaz de usuario: la técnica de grab-and-drop

En esta sección veremos los items más importantes en el diseño final de la interfaz de usuario, extendiendo los conceptos vistos en globalmente en el capítulo 6. Se sugiere el nombre de grab-and-drop como una extensión de la modalidad de drag-and-drop, la cual aparece en las GUIs.

Control del ambiente

El usuario compartirá la reunión virtual con los locutores, los cuales están distribuidos uniformemente, sentados en forma semicircular frente al él (fig 6.5). Acústicamente ellos serán presentados en el plano medio horizontal. Para las actividades de control el sistema debe proveer un *pointing device* o dispositivo de apuntamiento. En las dos versiones posibles estos pueden ser un joystick o un guante. Comentaremos como funciona esta última versión.

En una GUI tradicional, un usuario interactúa con las entidades apuntando y *clickeando* sobre ellas. También puede arrastrar, hacer doble click, clickear con el botón derecho del mouse , etc. En la interfaz propuesta, el usuario indica con la posición de la mano el objeto a activar, e indica la activación (o click) cerrando la mano. De esta forma básica el usuario puede seleccionar una entidad y generar el evento de activación asociado. En una sesión con el sistema HiperAudio, el usuario escuchará en un momento a un locutor. Al finalizar este, pueden aparecer links a otros locutores. Estos realizan un breve comentario. Al finallizar, el usuario puede cerrar la mano en la dirección del locutor deseado y así acceder al próximo nodo de información.

Iconos auditivos 3D

Hasta aquí se describió como funciona la activación de links, pero no se explicó como se llevan las tareas de control tal como backtracking, stop, repetición de tópicos, etc. La idea es simple: el usuario posee debajo del plano de los locutores una colección de íconos auditivos. Estos están en silencio mientras el usuario navega la conversación. Si en cierto momento el usuario precisa la funcionalidad de uno de ellos, cierra la mano en la posición del asistente, y este automáticamente los hace sonar. A este proceso lo denomino llamada de íconos o *icon calling*. El usuario reconoce la posición de cada uno de ellos, y puede cerrar la mano sobre el ícono deseado, recibiendo el feedback acústico adecuado. Una vez que lo tiene en la mano, lo puede arrastrar y soltar sobre la entidad destino de la operación deseada. Por ej. en caso de hacer backtarcking sobre un cierto locutor, toma el ícono asociado, lo arrastra y lo suelta sobre el locutor deseado. El diagrama de estados 7.11 muestra básicamente como es la modalidad de interacción del usuario con los íconos.

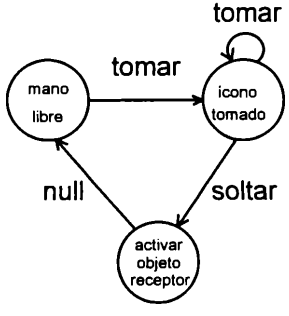


Fig. 7.11: La técnica de grab & drop usada en el HiperAudio queda reflejada en este diagrama de estados. Dependiendo del ícono tomado y el ícono destino, se activa de manera particular la funcionalidad que ofrece este último.

Esta técnica que en las GUI's se denomina *drag & drop*, es aplicada similarmente en este ambiente, permitiendo que en forma homogénea cada ícono representa una acción que será aplicada sobre la entidad receptora. Aquí se la denomina grab-and-drop.

De la misma manera que con el guante, la acción se puede llevar a cabo con un joystick. La ventaja obtenida es un menor actividad física la cual puede evitar cansancio pero el problema es que se pierde la ventaja de la modalidad cruzada kinestésica-acústica la cual refuerza el conocimiento de la ubicación de cada ícono como así el incremento de realismo del ambiente virtual presentado.

Personalización de la pantalla virtual acustica

El posicionamiento de los íconos en el ambiente debería ofrecer mínimo esfuerzo para las actividades más frecuentes. Además los íconos más usados deberían residir en reposo en las posiciones más cómodas para el usuario. Por estas razones, la primer tarea que realiza el usuario es una personalización o *customizing* del sistema. El ambiente espacial se descretiza en *voxels* o unidades elementales de volúmen, algo así como los pixels en una pantalla plana. Cerrando la mano en la posición deseada para cada ícono en particular, el usuario va determinando en que voxel reside cada ícono.

Cada ícono permanece en reposo hasta que el usuario hace un *icon calling*. En ese momento el usuario puede escucharlos en dos modalidades: simultánea o secuencial. En la primera todos los íconos suenan simultáneamente y en la segunda lo van haciendo en secuencia. La primera opción es conservativa en tiempo, útil cuando ya existe familiaridad con el sistema, pero puede resultar *a priori* confusa en la localización y discriminación. Esta posibilidad es válida pues los sonidos son 3D. La segunda opción permite un estudio más analítico de cada fuente, pero deseperdicia el recurso más caro que es el tiempo. El usuario puede seleccionar cualquiera de estas modalidades.

Elección de los sonidos icónicos

Cada sonido deberá representa una actividad o acción determinada. Las alternativas que mapean un cierta entidad o evento al dominio acústico sugeridas son:

• Iconos auditivos: sonidos cotidianos de la vida real [Gaver 86]

• Earcons: se refiere a una forma convencional de mapear eventos a simples sucesion de tonos, los cuales pueden ser combinados, transformados y heredados para crear una familia de earcons [Gaver 86]

• Sonidos genéricos: mapping metafórico entre eventos y eventos asociados extraídos de películas, shows de TV o situaciones convencionales [Cohen 93].

Cada una de estas estrategias posee ventajas y desventajas. Los íconos auditivos se apuntalan en el conocimiento del usuario acerca de los sonidos cotidianos y su significado, de esa manera pueden ser diseñados para ser obvios y fáciles de aprender. Sin embargo, el sonido apropiado a veces puede ser difícil de encontrar. Los earcons pueden ser sistemáticamente generados y diseñados para ser maximamente discriminables [Brewster 93]. Ellos, sin embargo, pueden ser difíciles de aprender y reconocer. Finalmente, los sonidos genéricos poseen un compromiso entre las dos alternativas anteriores: ellos pueden compartir la claridad de los íconos auditivos y la flexibilidad de los earcons, pero igualmente aparece la dificultad en el mapping evento-sonido tal como en los íconos auditivos y la dificultad de aprendizaje como en los earcons.

En general el uso del sonido en un sistema puede expresar globalmente estos tres tipos de actividades [Beadouin 94]

• *feedback:* que está haciendo el usuario;

• *notification:* que esta haciendo el sistema

• *awareness:* que están haciendo otras personas

En el sistema HiperAudio se utiliza la primer modalidad fundamentalmente: feedback para saber en que momento un ícono es tomado o soltado, o para saber cuan lejos está un ícono de la mano. Con respecto a la elección de sonidos se utilizaron íconos auditivos para el momento en que un ícono es tomado y un sonido puro musical para definir todo el instante en que el ícono es tomado. Esto genera dos fases de utilización de sonidos: uno sonido para tomar y soltarlo, y un sonido para el tiempo en que el ícono esta tomado. Este último sonido está procesado para generar un loop continuo sin sobresaltos bruscos. Según [Sikora 95] el uso intensivo de un sistema con señales acústicas es más agradable con earcons musicales, a pesar que son menos representativos que los iconos auditivos. Se han sugerido técnicas también para generar los sonidos icónicos en forma paramétrica, de tal manera de poder mapear ciertas variables del sistema a atributos del sonido [Gaver 92, Takala 92]

Otro tópico no tratado en la bibliografía es la distribución espacial de íconos auditivos, la cual es utilizada en HiperAudio. Por esa razón, el estudio de las modalidades adecuadas surgen de la prueba empírica de la interfaz. Actualmente se está trabajando para obtener los guidelines adecuados, siendos estos probablemente diferentes para usuarios ciegos y videntes.

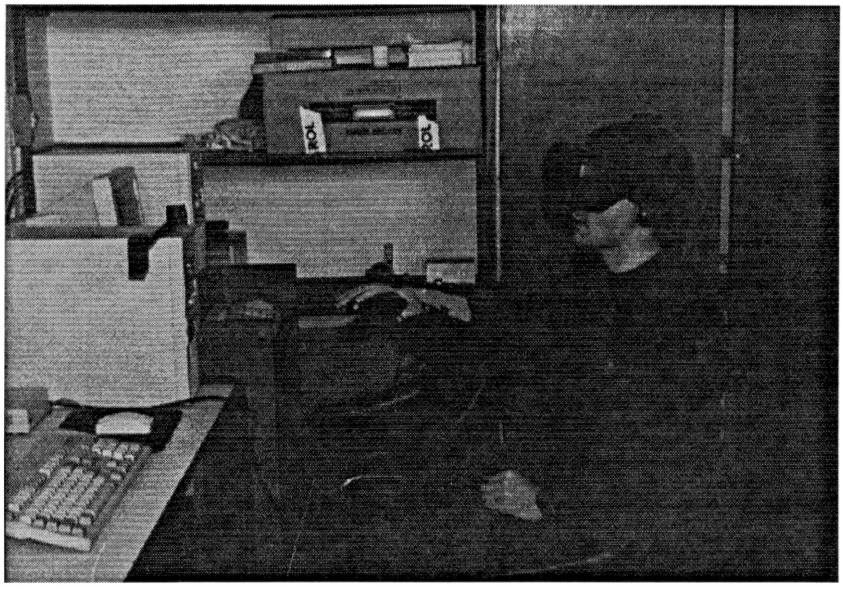


Fig. 7.12: Así es como se ve un usuario utilizando el HiperAudio usando el HMD y el guante. En este caso el usuario es el autor.

Conclusión

La implementación del HiperAudio permite explorar las virtudes del modelo hipermedial junto con las posibilidades de la tecnología de Realidad Virtual. La implementación del sistema implicó realizar un ambiente de autoría y un módulo de ejecución el cual permitió testear en sucesivos pasos de refinamiento, las ideas de interfaces planteadas en capítulos anteriores.

Nuevas técnica de browsing de información pueden potencialmente surgir de HiperAudio, el cual es el punto de partida para seguir testeando esta novedosa forma de obtener información.

Conclusión y Perspectivas Futuras

Capítulo

9

Resúmen

Navegar en el dominio temporal tal como lo plantea el sistema HiperAudio no es una tarea tan simple como lo es navegar espacial tal como lo hacen las aplicaciones que corren en una GUI. La naturaleza transitoria del sonido, no deja trazas de su existencia una vez que este concluye. Esta característica hace que la dificultad no surja solo de la navegación en sí misma, sino también de otras funcionalidades y características que surgen al utilizar una aplicación de naturaleza hipermedial. Comentaremos que aprendimos , cuales son los problemas y como se podría extender el sistema.

Lecciones aprendidas del HiperAudio: problemas, ventajas y extensiones del sistema.

La voz como medio portador de información

El oido no puede hacer browsing en el dominio temporal como lo hace la vista en el dominio espacial. Es por ello que HiperAudio a traves del uso e metáforas adecuadas y el uso del sonido 3D intenta resolver el problema impuesto por la naturaleza propia del sonido. La voz es efeciente para el que habla, pero usualmente pone una carga en el oyente [Grundin 88]. Lamentablemente un usuario ciego no posee muchas chances para revertir este problema. Escuchar es un proceso altamente consumidor de tiempo real de usuario.

Búsqueda de información

La búsqueda de informacion en cassettes de audio es dificil. HiperAudio ofrece una tecnología netamente superior a la que ofrecen los cassettes de audio, los cuales proveen un acceso secuencial a la informacion, pero existen varios factores que hacen que el browsing de audio inclusive usando HiperAudio sea confuso. Aunque hablar es mas rápido que tipear o escribir, oir algo hablado es mas lento que leer. Mas aún,el oido no puede hacer un overview de un cassette tal como lo hace la vista en una página de texto. Una grabación puede ser aumentada de velocidad para disminuir este problema, pero técnicas convencionales introducen un cambio en la altura del sonido, resultando en una pérdida de inteligibilidad. HiperAudio con el uso de caminos divergentes y con una extensión que provea acceso indexado intenta resolver estos problemas. En una extensión de HiperAudio se debería permitir viajar rápidamente por los nodos tal como lo hace una radio digital de automóvil. Casi todas las radios digitales de auto tienen botones de *scan* y *seek* . Scan es automático y simplemente va de una estación a otra. Seek permite a los usuarios viajar entre las estaciones pero bajo su control. El modo Scan en una radio puede ser frustrante debido a que es basado en intervalos fijos: después de unos segundos, la radio salta a la próxima estación sin tener en cuenta si es un aviso, la canción favorita o algo no deseado. Desde que el scan es un modo automático sin-manos, en el momento en que uno se da cuenta que algo interesante se esta pasando, la radio ha pasado a la próxima estación. El comando seek hace que la radio entre *en loop*, permitiendo que el usuario controle el período de escucha de cada estación, haciendo así mas deseable y eficiente la búsqueda. Estas sencillas lecciones podrían ser aplicadas a HiperAudio para recorrer ciertos *paths* predefinidos que siguen una línea de desarrollo, una idea o un conjunto de items de un locutor. De esta manera el usuario navegaría rápidamente por un hilo conductivo, parando y navegando a partir de allí a su propio ritmo.

Una extensión de HiperAudio debería incluír facilidades tal como sugiere [Arons 94]:

• *Skim* extraer rápidamente la informacion saliente de un chunk a modo de resúmen

• *Searching:* encontrar una pieza específica de información

• *Scanning:* similar al skim, pero da una connotación de una examinación rápida tal como la que efectúa la vista con una página de texto escrito

El sistema también debe proveer tal como lo hace SpeechSkimmer [Arons 94] la funcionalidad de búsqueda jerárquica: cuando buscamos información visualmente, nosotros tendemos a organizar nuestra búsqueda a lo largo del tiempo, buscando sucesivamente información con mas detalle. Por ejemplo, uno puede de un vistazo ir a la biblioteca, seleccionar el titulo adecuado, pasar páginas hasta encontrar el capítulo relevante, pasar los encabezamientos hasta encontrar la sección correcta, y alternativamente ir salteando y leyendo el texto hasta encontrar la información deseada. Para hacer browsing en segmentos de sonido, existiría una manera análoga: el usuario debe controlar el nivel de detalle, el régimen de *playback* y el estilo de presentación. Las metáfora del edificio sería un buen punto de partida para intentar proveer esta funcionalidad de jerarquía de tópicos..

HiperAudio como interfaz de acceso a WWW

Una extensión posible del sistema HiperAudio sería el acceso a WWW para usuarios ciegos. Es fácil crear un sintetizador de voz en el cual los fonemas están procesados para generar voz 3D. Con el standard HTML 3.0 es posible tipar los links, de tal manera de organizar por tipos la asignación de locutores a las páginas del Web. Si bien esta opción permite genericidad, pues el sintetizador no necesita de grabaciones previas de locutores y permite generación de voz *on the fly*, se pierde la característica de la voz hablada, la cual es un rico y expresivo medio [Chalfonte 91]. En adición al contenido léxico de las palabras pronunciadas, las emociones e importante información sintáctica y semántica es capturada por la altura, timing y la amplitud de la voz. En ciertos momentos, mas información semántica puede ser transmitida por el uso de silencios que por el uso de palabras!

Representación y expresión de la información

La navegación en el dominio de audio es mas difícil que en el dominio espacial. Conceptos tales como highlightning, a la derecha de , y selecciones basadas en

menúes son llevadas a cabo diferentemente que en interfaces visuales. Por ejemplo, uno no puede "clickear aqui", en el dominio de audio para obtener mas información: en el momento que la selección es hecha, el tiempo paso, y el "aquí" no existe mas. HiperAudio soluciona parcialmente esto, y con extensiones en el cual backgrounds de sonido y procesamiento con efectos, pueden dar las características de higlightning deseadas. La idea sería dar señales acústicamente identificables, manteniendo el feedback conciso con varios grados de feedback seleccionables. Un sistema de ayuda debería ser incluído para una interacción suave y eficiente. Con respecto al contenido propiamente del texto expresado por los locutores, éste debería respetar las máximas de Grice [Grice 75] acerca de como, que y cuando algo debe ser dicho. La base de datos hipermedial debe ser construida de tal maner de seguir un flujo conversacional suave en lugar de pasivamente ser paseado por menúes, prompts o despliegues continuos de íconos auditivos.

Autoría de un HiperAudio

La autoria de los sistemas de hipermedia es una de las tareas mas dificiles: los sistemas como HiperAudio tienen una complicación adicional debido a la naturaleza serial y no visual del sonido. Las señales grabadas no pueden ser manipuladas en el display de la misma manera como el texto o gráficos. Hay que tener en cuenta que las representaciones esquemáticas de las señales de audio pueden ser vistas en paralelo y manipuladas gráficamente, pero los segmentos de sonido representados en ese display no pueden ser escuchados simultáneamente. La retórica conversacional y la transcripción a texto escrito para una cabal comprensión del todo es una tarea larga y tediosa.

El recorte manual de grabaciones de audio produce segmentacion de alta calidad pero es dificil, lleva tiempo y es caro. La segmentación automática sería una solución. Segementadores automaticos de audio y descubridores de estructura interna inherente [Hawley 93] son esenciales para el éxito de sistemas tal como el presentado. "Encontrar la estructura" se refiere al descubrimiento de porciones importantes o enfatizadas de la grabación, es el equivalente a localizar párrafos o líneas delimitadoras de topicos en una página.

Compresión de la información

En este caso no nos referimos al almacenamiento del sonido, sino a la búsqueda de métodos que permitan ahorrar al usuario del sistema el tiempo de escucha. La

compresión de sonido es posible en diversos grados en función de la familiaridad del material por parte del usuario, el tamaño de cada segmento de voz, etc. La redundnacia temporal de la voz puede ser aprovechada para eleminar pausas innecesarias pero puede eliminar pistas sintácticas y semánticas. Métodos como la reproducción dicótica (un *sample* alternativamente a cada oído) han sido propuestos también [Orr 71].

Organización de la red de nodos y links

Una de los efectos percibidos en la prueba del HiperAudio es "... me estoy perdiendo algo importante de la conversación por haber elegido estos links...". Esta pregunta surge pues cada elección implica un compromiso entre la información que se obtiene y la que se deja de obtener. En el proceso de autoría se observa que es posible minimizar este efecto generando *una conversación para adelante.* Esto quiere decir que los links en general tiendan a cruzar opinión entre los locutores pero siempre con una continuidad argumental en el desarrollo de las ideas. De esta manera, si existe algo importante dado un contexto determinado, es altamente probable que esté próximo al nodo actual y con un link que permita su acceso. Esta modalidad de diseño surge naturalemente si la entrevista a cada locutor sigue una estructura similar, conservando un hilo argumental entre ellas.

Grab-and-drop de íconos auditoriales 3D

Esta novedosa técnica de presentación de "botones" en el espacio es otro aporte original de este trabajo. No existe literatura específica al respecto y éste puede ser un buen comienzo para estudiar la incorporación de esta modalidad a cualquier display virtual acústico. En particular, la falta de pistas visuales es lo que hace más atractiva la utilización de esta forma de elegir función a aplicar sobre una entidad objetivo. En una futura extensión en la cual la aplicación de HiperAudio sería la generación de material de entretenimiento, el grado de inmersión, interactividad y actividad física del usuario provisto por el ambiente kinestésico-auditivo no tendría rival comparado con las técnicas de input tal como las que ofrece el teclado o el mouse. El rendering de los objetos y la interacción provista son dos factores fundamentales en un juego estilo *arcade.* Piense lo que sería jugar al DOOM en modo texto!

Conclusión

Interfaces acústicas puras recién comienzan a ser discutidas entre los investigadores de la interacción hombre-máquina. Generalmente el sonido viene

a complementar o extender lo que una interfaz gráfica ofrece. Como dijimos a lo largo de este trabajo, un usuario no vidente impone una perspectiva de diseño diferente a la mencionada. Es por ello que el sistema HiperAudio intenta crear un framework apto para la interacción hipermedial para impedidos visuales. La tarea de implementación de estas ideas no fue trivial y demandó casi dos años de trabajo. El valor final del producto puede ser discutido y actualmente comienza a ser testeado por usuarios ciegos. Pero en lo que no hay duda es que el backgroud generado es un importante legado para la futura investigación de los tópicos propuestos.

Como resúmen, el trabajo original de esta tesis conceptualmente se centró en:

- El uso del modelo hipermedial como método de acceso y presentación de información a no videntes
- El uso del sonido 3D como portador de información y como medio subyacente para permitir recrear virtualmente la metáfora de la conversación espacial
- La idea de utilización de íconos auditoriales 3D y de la aplicación de la técnica de grab-and-drop como modalidad de interacción en un ambiente sin pistas visuales

En lo que refiere a lo implementativo:

- Un programa para el procesamiento de sonido 3D *off-line* , con el objeto de generar sonido 3D a partir de una fuente monofónica cualquiera
- Una herramienta para la edición de un HiperAudio
- Un módulo para la ejecución en tiempo real de una instancia particular de HiperAudio

Aunque todos los tópicos son explorados en el contexto de acceso a información hipermedial para no videntes, los resultados son aplicables a otras interfaces acústicas puras también como a interfaces visuales que empleen íconos con feedback auditivo.

No quedan dudas que mejores interfaces para discapacitados, inexorablemente implicará mejores interfaces para todos.

Glosario

ADC	Analog to Digital Converter
anchor	Punto de partida o llegada de un link en una aplicación hipermedial
chunk	Segmento o porción de información atómica
DA	Display Acústico
DAC	Digital to Analog Converter
data-driven	Arquitectura de software o hardaware en donde el control es manejado por el flujo o acceso a los datos
DAV	Display Acústico Virtual
DFT	Discrete Fourier Transform: Transformada discreta de Fourier
drag-and-drop	Técnica usual en GUIs, en la cual un objeto es seleccionado con el mouse, arrastrado y soltado sobre otro objeto gráfico, con el fin de realizar una acción que involucra el objeto

	manipulado y el objeto destino
DSP	Digital Signal Processing: Procesamiento digital de señales
FFT	Fast Fourier Transform: Transformada rápida de Fourier
gestalt	Rama de la sicología que estudia la propensión que posee el ser humano para reconocer patrones y configuraciones que aparecen en el ambiente. "Gestalt" en alemán, significa una entidad independiente la cual tiene una forma definida.
grab-and-drop	Una extensión propuesta en esta tesis de la técnica de drag-and-drop de las GUIs. La diferencia aquí se centra en el uso de íconos auditoriales 3D como entidad representativa de la interacción. *A priori* no posee pistas visuales este tipo de interacción.
GUI	Graphical User Interface: Interface gráfica de usuario
hearing	"hearing", del inglés, sentido del oído o acción de escuchar.. En este trabajo se utiliza como sinónimo la palabra *percepción*, siendo ella mas adecuada que la traducción textual
HMD	Head Mounted Display: Display montado en la cabeza, o más intuitivamente, el casco de las aplicaciones de realidad virtual
HRTF	Head Related Transform Function: Función de transferencia relativa de la cabeza, es el efecto particular que impone el canal auditivo y la cabeza a una señal que llega al oyente, afectando la amplitud y fase relativa de cada componente del sonido en cuestión

HTML	HyperText Markup Language: un standard para describir contenido hipertextual independiente de la plataforma de presentación.
IID	Interaural Intensity Difference: se refiere a la diferencia de amplitud con que una señal llega a cada uno de los oídos
ITD	Interaural Time Difference: se refiere a la diferencia en el tiempo con que una señal llega un oído con respecto al el otro
layout	Descripción o disposición de objetos sobre un medio, generalmente plano
mS	miliSegundo
nonspeech	Señal acústica que no es voz, es decir un ruido, el sonido de un instrumento musical o un tono.
on-the-fly	Computado en tiempo real, en el momento que se ejecuta
PAL	Phase Alternate Line: un sistema de codificación de señal de video
pitch, roll , yaw	Términos que definen rotación, elevación de la nariz e inclinación de la cabeza
sampler	Términos para referirse a un dispositivo que digitaliza/reproduce señales, generalmente utilizado como instrumento musical.
samples	Valores obtenidos al digitalizar una señal analógica
sampling	Acción de tomar muestras digitales de una señal analógica
stream segregation	La capacidad de atender a un grupo de sonidos y asociarlos a una sola entidad

	conceptual
VAD	Virtual Acoustic Display: un medio preciso para expresar acústicamente en un ambiente virtual a diferentes entidades
VR	Virtual Reality: realidad virtual
WWW	World Wide Web: conjunto de computadoras que operan bajo Internet y que ofrecen acceso a documentos organizados de manera hipermedial

Referencias

[Alpha 95] Alphatron User's Manual, Crystal River Engineering Inc., 490 California Ave, Suite 200, Palo Alto, CA 94306, 1995

[Arons 91] Arons B., Hyperspeech: Navigating in speech-only hypermedia, in Proceedings of Hypertext '91, ACM, 1991, pp. 133-146

[Arons 93] Arons B., SpeechSkimmer: Interactively Skimming Recorded Speech, In Proceedings of the ACM Symposium on User Interface Software and Technology , UIST, ACM Press, Nov. 1993

[Arons 94] Arons B., Interactively Skimming Recorded Speech, Ph.D Thesis, MIT, Febrero 1994.

[Beadouin 94] Beadouin-L M., Gaver W., ENO: Synthesizing Structured Sound Spaces, UIST '94, ACM Conference on User Interface System Tailoring, November 1994, pp.49-57

[Begault 94] Begault D., 3D Sound for Virtual Reality and Multimedia,

AP Professional, Cambride, MA, 1994

[Belkin 87] Belkin N., Croft W., Retrieval Techniques, in Williams ME Editor, Annual Review of Information Science and Technology (ARIST), 22, pp. 109-145

[Blenkhorn 92]. Blenkhorn P., Requirements for screen access software using synthetic speech, in Zagler W. editor, Computer for Handicapped Persons, Oldenburg: Wien, pp. 31-37

[Bregman 94] Bregman A., Auditory Scene Analysis: The perceptual Organization of Sound, MIT Press, 1994

[Brewster 93] Brewster S., Wright P., Edwards A., An evaluation of earcons for use in auditory human-computer interfaces, ACM Conference on Human factors in Computing Systems, INTERCHI '93, Amsterdam, The Netherlands, April 24-29, 1993, pp. 222-227

[Brondmo 89] Brondmo H., Davenport G., Creating and Viewing the Elastic Charles - A Hypermedia Journal, Hypertext II, Jun, York, UK

[Buchanan 92] Buchanan M.C., Zellweger P.T, Specifying Temporal Behavior in Hypermedia Documents, in ACM ECHT 92 Conference, Milan, Italia, Noviembre. 1992 ,pp.262-271

[Burgess 92] Burgess D., Verlinden J., An architecture for spatial audio servers, Tech Report GIT-GVU-92-24, Georgia Tech

[Canter 77] Canter D., "The psychology of place", London: Architectural Press, 1977

[Cohen 93] Cohen J., "Kirk here:" Using genre sounds to monitor background activity. Adjunt Proceedings, ACM Conference on Human factors in Computing Systems (INTERCHI '93, Amsterdam, The Netherlands, April 24-29, 1993), pp. 63-64

[Conklin 87]. Conklin J., Hypertext: An Introduction and Survey, Computer, September 1987, pp. 17-41

[Crouch 89]. Crouch D., Crouch C., Andreas G., The Use of Cluster Hierarchies in

Hypertext Information Retrieval, Proceedings of Hypertext '89, pp. 269-292

[Chalfonte 91] Chalfonte B., Fish R., Kraut R., Expressive Richness: A Comparison of Speech and Text as Media for Revision. En Proceedings of CHI (New Orleans, LA, Apr. 28-May 2), ACM, New York, 1991, pp. 21-26

[Erickson 90] Erickson T., Working with Interface Metaphors, in The Art of Human-Computer Interface Design, Addison-Wesley Publishing, pp. 65-73

[Fairchild 88]. Fairchild K., Poltrock S., Furnas G., Sem Net: 3D Graphics Representations of Large Knowledge Bases, in Cognitive Science and Its Applications for HCI, Lawrence Erlbaum Associates, Hillsdale, NJ

[Feiner 88] Feiner S., Seeing Forest for the Trees: Hierarchical Display of Hypertext Structure, Proceedings of Conference on Office Information Systems, ACM, pp. 205-212

[Gaver 86] Gaver W., Auditory icons: Using sound in computer interfaces. Human-Computer Interaction, 2 (2), pp. 167-177, 1986

[Gaver 92] Gaver W., Using and creating auditory icons, en ICAD 92, Auditory Display: Sonification, Audification and Auditory Interfaces, pp. 417-446, Addison-Wesley, 1994

[Gehring 90] Gehring B., Focal Point 3D Sound User's Manual, Gehring Research Corporation, 189 Madison Avenue, Toronto, Ontario, Canada, M5R 2S6

[Gibson 79] Gibson, J.J., The Ecological Approach to Visual Perception. Boston: Houghton Mifflin, 1979.

[Gierlich 92] Gierlich H., The application of Binaural Technology, Applied Acoustic, 36, 219-243

[Glushko 89] Glushko R., Design Issues for Multi-Documents Hypertext, Proceedings of Hypertext '89, November 1989, pp. 51-60

[Grice 75] Grice H., Logic and Conversation. In Speech Acts, edited by P. Cole and J. Morgan. Syntax and Semantics, vol. 3. New York:Academic

Press, 1975, pp. 41-58

[Grundin 88] Grundin J., Why CSCW Application Fail: Problems in the design and Evaluation of Organizational Interfaces. En Proceedings of CSCW (Portland, OR, Sep. 26-28), ACM, New York, 1988, pp.85-93

[Halasz 87] Halasz F., Morgan T., Trigg R., Notecards in a Nutshell, Proceedings ACM CHI 87, Toronto, Canada, pp. 45-52

[Hatwell 93] Hatwell Y.,"Images and non-visual spatial representations in the blind", in Proceedings of the INSERM-SETAA conference Non-Visual HCI, Paris, March 1993 , pp. 13-34

[Hawley 93] Hawley M., Structure out of sound, Ph. D dissertation, MIT, Sept. 1993

[Hinckley 94] Hinckley K., Paush R., Goble J., Kassell N., A Survey of Design Issues in Spatial Input, in UIST'94 ACM Symposium on User Interface Software and Technology, Marina del Rey, California, USA, November 2-4, 1994, pp. 213-222

[Hirata 93] Hirata K., Hara Y., Shibata N., Hirabayashi F., Media Based Navigation for Hypermedia Systems, in proceedings of Hypertext '93, ACM Conference in Hypertext Technology, pp. 159-173, Seattle USA

[Kawalski 93] Kawalski R., The Science of Virtual Reality & Virtual Environments, Wokingham, England: Addison-Wesley

[Kimber 95] Kimber D., Wilcox L., Chen R., Moran T., Speaker Segmentation for browsing recorded audio, in companion proceedings of CHI 95

[Kramer 94] Kramer G., An introduction to Auditory Displays, en ICAD 92, Auditory Display: Sonification, Audification and Auditory Interfaces, pp. 1-78, Addison-Wesley, 1994

[Landau 88] Landau B., The construction and use of spatial knowledge in the blind and sighted children, in Spatial Cognition:Brain Bases and development, Lawrence Erlbaum Editors, New Jersey:1988, pp.343-371

[Lumbreras 93a] Lumbreras M., A hypertext for blind people, poster of Hypertext 93,

ACM Confrence on Hypertext Technology, Seattle, Nov 1993, USA

[Lumbreras 95a] Lumbreras M., Rossi G., A metaphor for the visually impaired: browsing information in a 3D auditory environment, in Proceedings Companion CHI 95, ACM Conference on Human Interaction, Denver, Colorado, Mayo 1995, pp.261-262

[Lumbreras 95b] Lumbreras M., Barcia M., Hyperaudio: un ambiente hipermedial para ciegos, 24 JAIIO, 7-9 Agosto de 1995, Ciudad Universitaria, pp. 6.77-6.88

[Monk 89] Monk A., Getting to a Known Locations in a Hypertext, Hypertext II, Jun, York, UK

[Muller 92] Muller M., Farrel R., Cebulka K., Smith J., Issues in the Usability of Time-Varying Multimedia, Multimedia Design, ACM Press, 1992, pp. 7-38

[Mynatt 92]. Mynatt E., Edwards W., The Mercator Environment. A Non Visual Interface to X- Windows and Unix Workstation. GVU Tech Report GIT-GVU-92-05, Feb 1992

[Oppenheim 89] Oppenheim A.V., Schafer R.W., Discrete-time Signal Processing, Englewood Cliffs, NJ:Prentice Hall

[Orr 71] Orr D., A perspective on the percption of time compressed speech, en Perception of Language, editado por P. Kjeldergaard, D. Horton y J. Jenkins, Charles Merril Publishing Company, 1971, pp. 108-119

[Pohlman 91] Pohlman K., Advanced Digital Audio, SAMS Books, 1991

[Preece 94] Preece J., Human Computer Interaction, Addisson Wesley, 1994

[Rayleigh 07] Rayleigh L., Sttrutt J., On our perception of sound direction, Philosophy Magazine, 13, pp. 214-232, 1907

[Rossi 93] Rossi G., Lumbreras M., Zato G., On the use of multimedia as a rehabilitation/integration technology, ECART 2, European Conference on Advancement of Rehabilitation Technology, 26-28 Mayo de 1993, Estocolmo, Suecia, pp.22.4

[Scaletti 92] Scaletti C., Sound Synthesis Algorithms for Auditory Data Representations, en ICAD 92, Auditory Display: Sonification, Audification and Auditory Interfaces, pp. 223-252, Addison-Wesley, 1994

[Sikora 95] Sikora C., Roberts L., Murray L., Musical vs. real feedback signals, in companion proceedings of CHI 95

[Stiefelman 93] Stiefelman L., Arons B., Schmandt C., Hulteen E.,Voice Notes: A Speech Interface for a Hand-Held Voice Notetaker, in Proceedings of INTERCHI '93, ACM, April 1993,pp.179-186

[Takala 92] Takala T., Hahn J., Sound Rendering, Computer Graphics,26, 2 July 1992, pp. 211-219

[Vanderheiden 92] Vanderheiden G., A White paper on the design of software application programs to increase their accesability for people with disabilities, Madison: University of Winsconsin-Madison, Trace R&D Center, 1992

[Weber 93] Weber G., Kochanek D. et al, Access by blind people to interaction objects in MS Windows, in Proc. ECART 2, European Conference on the Advancement of Rehabilitation Technology , Stockholm, Mayo 1993, pp.2.2

[Wenzel 92] Wenzel E.M., Localization in Virtual Acoustic Displays, Presence, vol.1 number 1, 1992, pp. 80-017

[Wheeler 93] Wheeler A., Ellinger J., Glicker S., The Design and Implementation of an Experimental Virtual Acoustic Display, Technical Report GR-EM-93-1, February 1993, Electrical and Computer Engineering Department, Univeristy of Texas at Austin

[Wightman 89] Wightman F.L., Kistler D.J., Headphone Simulation of Free Field Listening: Stimulus Synthesis, Journal of the Acoustical Society of America , 85, pp. 858-867

[Yankelovich 95] Yankelovich N., Levow G., Marx M., Designing SpeechActs: Issues in Speech User Interfaces, in Proceedings of Human Factors in Computing Systems, CHI'95, ACM, Denver, Colorado, Mayo 1995, pp. 369-376