

UNIVERSIDAD POLITÉCNICA DE VALENCIA

DEPARTAMENTO DE SISTEMAS INFORMÁTICOS Y COMPUTACIÓN



MÁSTER UNIVERSITARIO EN INTELIGENCIA ARTIFICIAL,
RECONOCIMIENTO DE FORMAS E IMAGEN DIGITAL

DESARROLLO DE UN SISTEMA DE
MULTICONFERENCIA INMERSIVA CON
AUDIO 3D PARA MÓVILES

TESINA DE MÁSTER

Presentado por:
Emanuel Aguilera Martí

Dirigido por:
Dr. José Javier López Monfort

Valencia, 2011

A mi padres, Rubén y Mónica, y a mi hermana, Anaí, por apoyarme incondicionalmente, tenerme paciencia y por lo mucho que les quiero.

Índice General

Resumen.....	4
Agradecimientos.....	6
1. Introducción	7
1.1. Contexto y motivación del proyecto	7
1.2. Multiconferencia y telefonía móvil	8
1.3. Inteligibilidad y <i>Cocktail Party Effect</i>	9
1.4. Objetivos del proyecto	10
1.5. Estructura de la memoria.....	13
2. Estado del arte de la multiconferencia y los terminales móviles.....	15
2.1. Evolución de la telefonía móvil	15
2.2. Terminales móviles inteligentes	17
2.2.1. Introducción	17
2.2.2. Sistemas operativos para móviles	18
2.2.3. <i>Tablets</i>	20
2.3. Soluciones comerciales de multiconferencia por móvil.....	21
2.4. Conclusiones.....	22
3. Estudio de los requisitos del proyecto	24
3.1. Requisitos y plataformas.....	24
3.2. Experiencia de uso para el cliente del servicio.....	25
3.3. Gestión del demostrador.....	26
3.4. Aspectos técnicos y limitaciones	27
4. Tecnología utilizada.....	30
4.1. Introducción	30
4.2. Códec de compresión Speex	31

4.3. Sistema binaural HRTF de espacialización de sonido	32
4.4. Sistemas operativos para móviles.....	35
4.4.1. iOS.....	35
4.4.2. Android.....	36
5. Desarrollo de las aplicaciones del sistema	38
5.1. Introducción	38
5.2. Diseño de la interfaz gráfica.....	39
5.2.1. Homogeneidad entre los sistemas operativos.....	39
5.2.2. Estructura de la interfaz gráfica	40
5.2.3. Arte de la interfaz gráfica.....	42
5.3. Optimización de los algoritmos y tecnologías.....	43
5.4. Implementación de la aplicación para los terminales	45
5.5. Implementación del servidor de multiconferencia.....	48
6. Resultados y trabajo futuro.....	50
6.1. Resultados obtenidos	50
6.2. Repercusión del proyecto	51
6.2.1. Mobile World Congress 2011 Barcelona.....	52
6.2.2. SICARM 2011 Murcia.....	53
6.2.3. Repercusión en la prensa.....	53
6.3. Trabajo futuro	54
6.3.1. Mejoras del demostrador.....	54
6.3.2. Posibilidades de comercialización.....	55
Bibliografía.....	56

Índice de figuras

Figura 1. Cocktail Party Effect.	10
Figura 2. Percepción espacial del usuario con audio 3D.....	11
Figura 3. Arquitectura cliente-servidor para multiconferencia con diferentes terminales móviles.	12
Figura 4. Concepto básico de la interfaz táctil de usuario para mover intuitivamente a los participantes por la sala.....	13
Figura 5. Teléfono móvil inteligente del fabricante HTC.....	17
Figura 6. La <i>tablet</i> iPad 2 de Apple.....	20
Figura 7. Interfaz táctil para mover a los participantes.....	25
Figura 8. Diferente respuesta al impulso según el camino entre la fuente y cada oído.....	34
Figura 9. Pantalla principal del sistema operativo iOS.....	36
Figura 10. Pantalla principal de Android en un terminal HTC.....	37
Figura 11. Boceto de la pantalla principal.....	40
Figura 12. Boceto de la pantalla inicial.....	41
Figura 13. Estructura de pantallas de la aplicación para el móvil.....	41
Figura 14. Icono principal de la aplicación.....	43
Figura 15. Pantallas de inicio, principal y diferentes menús en iOS y Android respectivamente.	47
Figura 16. Pantalla principal del servidor de multiconferencia.	49
Figura 17. Autoridades probando la “Multiconferencia Inmersiva” en el MWC 2011.	52

Resumen

La telefonía móvil y el uso de terminales móviles avanzados está convirtiéndose en la forma de comunicación a distancia entre personas más importante en nuestros días, desplazando a la telefonía fija y a otros medios de comunicación.

La multiconferencia, como forma de mantener reuniones entre más de dos personas distantes entre sí, proporciona unas ventajas incuestionables y ya se viene utilizando de manera habitual en la telefonía fija.

De la misma forma, un terminal móvil también puede utilizar estos servicios de multiconferencia que proporcionan las operadoras, como lo hace un terminal de telefonía fija. Sin embargo, los terminales móviles de hoy en día disponen de muchas más prestaciones que los terminales fijos, que han quedado relegados prácticamente a la pura comunicación vocal.

La idea de aprovechar la potencia de los terminales móviles para realizar multiconferencias más realistas, adquiere en nuestros días un gran interés comercial. Una de estas mejoras está relacionada con la incorporación del sonido espacial a la multiconferencia.

Está comprobado que el sistema auditivo humano, que es por naturaleza binaural, dispone de una capacidad especial para localizar a los sujetos en la escena sonora mediante la combinación de las señales de los dos oídos. Pero además, dispone también de capacidad para separar y comprender mejor a los locutores cuando estos hablan a la vez o se interrumpen. Esta capacidad se denomina *cocktail party effect* y ayuda a mejorar aspectos como la inteligibilidad, la inmunidad a interferencias y la comprensión del habla.

El hecho de que hoy en día la mayoría de terminales móviles dispongan de salida de audio estéreo para ser utilizada con auriculares, así como gran capacidad de cálculo para trabajar con algoritmos de procesado de audio digital, abre la posibilidad de mejora en la multiconferencia incorporando el sonido espacial a la misma en los terminales móviles.

Esta tesina consiste en el desarrollo de una plataforma piloto de evaluación de un sistema de multiconferencia con audio espacial para terminales móviles (iOS y Android). Este proyecto ha sido llevado a cabo por el presente autor, bajo la dirección del tutor de esta tesina, desde el instituto de investigación iTEAM de la UPV para la empresa Telefónica. Dentro de su estrategia de innovación en servicios, dicha empresa solicitó al iTEAM el desarrollo de este sistema para evaluar el posible interés del mercado en estas tecnologías.

De esta forma se ha implementado un sistema donde cada usuario, a través de su teléfono móvil inteligente, puede escuchar al resto de participantes de la multiconferencia como si estuvieran realmente en distintas posiciones de una sala virtual, mejorando sustancialmente la identificación e inteligibilidad de los locutores. Mediante la pantalla táctil, cada usuario tiene la libertad de colocar al resto en las posiciones virtuales que quiera, a la vez que disfruta de una sensación de presencia realista e impactante en la comunicación gracias al audio espacial.

A lo largo de esta memoria se van a explicar las bases en que se sustenta el proyecto, cómo se ha desarrollado y los resultados obtenidos. Se intentará abarcar todos los aspectos importantes de un proyecto así, que sin duda serán de gran interés para el lector, aunque por motivos obvios de confidencialidad se han tenido que omitir los detalles más técnicos de muchos de los desarrollos implementados.

Agradecimientos

Quisiera expresar mi más sincero agradecimiento a:

 Mi jefe en el trabajo y tutor en esta tesina, José Javier, por guiarme y enseñarme el sentido de la ingeniería.

 A mis compañeros de laboratorio del GTAC (iTEAM) por crear un ambiente de trabajo tan ameno y agradable cada día.

 A Alma y Ana, de Telefónica, por la oportunidad de realizar este proyecto y las agradables reuniones que hemos mantenido.

Capítulo 1

Introducción

1.1. Contexto y motivación del proyecto

Este proyecto se enmarca dentro del contrato de I+D que la empresa Telefónica suscribe con el instituto de investigación iTEAM de la Universidad Politécnica de Valencia para el desarrollo de una plataforma piloto de evaluación de un sistema de multiconferencia con audio espacial para terminales móviles.

Para Telefónica, la innovación es un pilar fundamental de su estrategia para cumplir con los requisitos más exigentes de sus clientes. En este ámbito de innovación, se plantea la prueba de concepto de la “multiconferencia inmersiva” o multiconferencia con sonido 3D.

Con esta prueba tecnológica se busca mejorar la calidad de la voz en las multiconferencias móviles así como aumentar el nivel de comprensión, permitiendo identificar el participante que está hablando en cada momento mediante el uso de técnicas de audio espacial.

Todos sabemos que la comprensión de quién habla en una multiconferencia es más complicada cuantos más participantes se encuentren conectados. Está demostrado que las personas distinguen y comprenden mejor a varios locutores si la voz de cada uno se escucha proviniendo de una dirección distinta. Aprovechando esta capacidad del oído humano, con esta aplicación cada participante tendrá la sensación de encontrarse inmerso en una sala virtual con el resto de participantes, pudiendo colocarlos en cualquier posición. Esta sensación de realismo donde cada participante parece estar donde se le ha colocado es asombrosa.

Aunque este proyecto se trata únicamente de un sistema piloto, Telefónica ve numerosas posibilidades para lanzar un servicio de multiconferencia de esta calidad al mercado, tanto en el entorno empresarial (multiconferencias de trabajo), como en el ámbito residencial (con amigos, familia,...).

Telefónica S.A. es una de las operadoras de telecomunicaciones líderes a nivel mundial que ofrece servicios de telefonía fija, telefonía móvil y ADSL. Como multinacional tiene su sede en Madrid (España) y presencia en gran parte de Europa y Latinoamérica, ocupando la quinta posición por capitalización bursátil a nivel mundial en el sector de telecomunicaciones. En junio de 2011 su número total de clientes en todo el mundo era de 295 millones. En España opera sus productos bajo la marca comercial Movistar dando servicio a 47 millones de clientes [1].

El iTEAM (Instituto de Telecomunicaciones y Aplicaciones Multimedia) es un centro de investigación integrado en la Ciudad Politécnica de la Innovación de la Universidad Politécnica de Valencia, donde desarrolla sus actividades de I+D+i dentro del área de las Tecnologías de la Información y las Comunicaciones. El iTEAM está formado por 9 grupos de investigación que abarcan todas las áreas científicas relacionadas con la Ingeniería de las Telecomunicaciones y que cuentan con más de 90 investigadores en total [2].

Dentro del iTEAM, el Grupo de Tratamiento de Audio y Comunicaciones (GTAC), al que el presente autor pertenece, abarca el área de trabajo dedicada a soluciones de procesamiento de señal para aplicaciones de audio, multimedia y sistemas de comunicaciones digitales [3].

1.2. Multiconferencia y telefonía móvil

La telefonía móvil se ha convertido en la forma de comunicación a distancia entre personas más importante en nuestros días, desplazando a la telefonía fija y a otros medios de comunicación. Así mismo, la nueva generación de terminales móviles avanzados o “inteligentes” está incorporando nuevas variantes en la comunicación móvil entre las personas, con nuevas aplicaciones de videoconferencia, de mensajería instantánea, de redes sociales, etc.

Por otra parte, los servicios de voz que permiten reunir en una misma llamada a más de dos personas proporcionan unas ventajas incuestionables hoy en día. Este tipo de servicios suelen ser llamados teleconferencia, audioconferencia, *conference call* o multiconferencia, como lo llamaremos de aquí en adelante. Las multiconferencias son muy útiles sobre todo en el ámbito empresarial porque ahorran muchos costes en desplazamientos físicos. Permiten mantener reuniones

entre distintas sedes de una empresa para coordinar proyectos, realizar seguimientos, comunicar estrategias, así como reuniones con clientes y proveedores sin los inconvenientes de los desplazamientos. Es por esto, que ya desde hace mucho tiempo, las empresas las vienen utilizando de manera habitual en la telefonía fija a través de los distintos servicios para empresa que ofrecen las operadoras.

De la misma forma, un terminal móvil también puede utilizar estos servicios de multiconferencia que proporcionan las operadoras, como lo hace un terminal de telefonía fija. Esto aporta obviamente una primera ventaja, ya que el usuario empresarial no tiene porqué estar en su oficina si no que puede mantener una reunión desde donde se encuentre en cualquier momento, ya sea en la calle, un aeropuerto, un hotel o realizando labores de campo. Así mismo, con un sistema de acceso y tarifas adecuadas, el servicio puede ser atractivo para los usuarios domésticos (no empresariales) que podrían realizar reuniones de audio con sus amigos y familiares desde cualquier sitio.

Pero además, los terminales móviles de hoy en día disponen de muchas más prestaciones que los terminales fijos, que han quedado relegados prácticamente a la pura comunicación vocal. La idea de aprovechar la potencia de los nuevos terminales móviles para realizar multiconferencias mucho más realistas es una segunda y muy importante ventaja que adquiere en nuestros días un gran interés comercial. Una de estas mejoras está relacionada con la incorporación del sonido espacial a la multiconferencia.

1.3. Inteligibilidad y *Cocktail Party Effect*

Cuando se observa la experiencia del usuario en los actuales sistemas de multiconferencia, siempre nos encontramos en mayor o menor medida con problemas derivados del carácter monofónico del servicio. Es decir, los sistemas actuales mezclan las voces de todos los interlocutores en una única señal de audio que se reproduce por un único altavoz en cada terminal. Esto no solo hace poco realista la conferencia sino que a menudo dificulta la inteligibilidad en la conversación cuando habla más de una persona a la vez porque se superponen sin más las voces. Podemos añadir el problema de distinguir quién está hablando en cada momento, sobre todo cuando no se conocen bien las voces o a los interlocutores, algo que sucede más en las comunicaciones empresariales.

Por otra parte, está comprobado que el sistema auditivo humano, que es por naturaleza binaural, dispone de una capacidad especial para localizar a los sujetos en la escena sonora mediante la combinación de las señales de los dos oídos. Pero además, dispone también de capacidad para separar y comprender mejor a los

locutores cuando estos hablan a la vez o se interrumpen. Esta capacidad se denomina *cocktail party effect* y ayuda a mejorar aspectos como la inteligibilidad, la inmunidad a interferencias y la comprensión del habla. Un ejemplo clásico de esta habilidad que tenemos es el hecho de que cuando nos encontramos en una sala ruidosa y llena de gente, aún así podemos escuchar y entender a la persona con la que estamos hablando, mientras ignoramos las otras conversaciones y el ruido de fondo [4].



Figura 1. Cocktail Party Effect.

El hecho de que hoy en día la mayoría de terminales móviles dispongan de salida de audio estéreo para ser utilizada con auriculares, así como capacidad de proceso para trabajar con algoritmos de audio digital, abre la posibilidad de mejora en la multiconferencia incorporando el sonido espacial a la misma en los terminales móviles. De esta forma se pueden explotar las capacidades naturales que tenemos para focalizar nuestra escucha en una determinada dirección, que como se ha comentado antes, son habilidades basadas sobre todo en un efecto binaural.

1.4. Objetivos del proyecto

El objetivo de este proyecto es el desarrollo de una plataforma piloto de evaluación de un sistema de multiconferencia con audio espacial para terminales móviles (iOS y Android). Como se verá más adelante, la parte más innovadora del proyecto es la inclusión del **audio espacial en la multiconferencia**, algo que todavía no se ha visto en ninguna solución comercial. Así mismo esto va acompañado del diseño y desarrollo de una **novedosa interfaz gráfica y táctil** que aumentará la usabilidad al sistema.

El sistema permitirá a un grupo de usuarios de terminales móviles establecer una multiconferencia, donde utilizarán los auriculares estéreo del terminal móvil (aunque el sonido va más allá de los *panning* estéreo tradicionales). Cada uno de los usuarios oirá al resto de participantes formando parte de una escena sonora, escuchando a cada uno de ellos en una posición espacial distinta, de forma similar a como si se encontrasen sentados alrededor de una mesa de reuniones virtual.

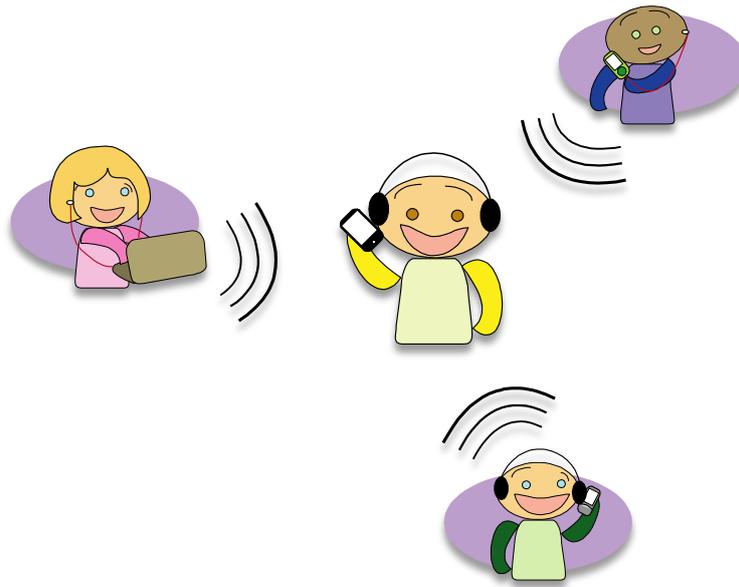


Figura 2. Percepción espacial del usuario con audio 3D.

De esta forma, se obtienen varias ventajas:

- **El participante asocia inmediatamente una posición espacial con una persona.** Esto le ayuda a identificar quién está hablando en cada momento, aunque no conozca previamente sus voces.
- **Aumenta la sensación de presencia.** La conversación es más real y da mayor sensación de estar en la reunión cara a cara.
- **Se mejora la inteligibilidad en caso de hablar varios a la vez.** Es mucho más fácil entender a los interlocutores en los momentos en que se interrumpen o hablan a la vez porque se puede prestar atención a determinadas direcciones espaciales.

A grandes rasgos, el sistema tendrá una arquitectura cliente-servidor (ver figura 3) donde se deberán desarrollar las siguientes funciones:

- Por parte del **servidor** deberán:
 - **Gestionar la multiconferencia**, controlando la incorporación o salida de los usuarios.
 - Distribuir las señales de voz a todos los participantes.
- Por parte de los **terminales móviles** deberán:
 - **Comprimir la señal de voz** que capturan su micrófonos y enviarla al servidor, y **descomprimir** las señales del resto de participantes que reciben del servidor.
 - **Sintetizar la escena de audio espacial** a partir del audio recibido y la información de la posición de cada participante.
 - Ofrecer una **interfaz visual y táctil**, cómoda e intuitiva, para que cada usuario sitúe al resto en la posición de la escena sonora que desee (ver figura 4).

Por tanto, el trabajo final a realizar en esta tesina se puede resumir en:

- Desarrollo del **software del servidor** central de multiconferencia.
- Desarrollo y programación de las **técnicas de codificación** de la voz para los terminales móviles.
- Desarrollo y programación de los **algoritmos de espacialización del sonido** para los terminales móviles.
- Desarrollo del **software de interfaz de usuario** para los terminales móviles.

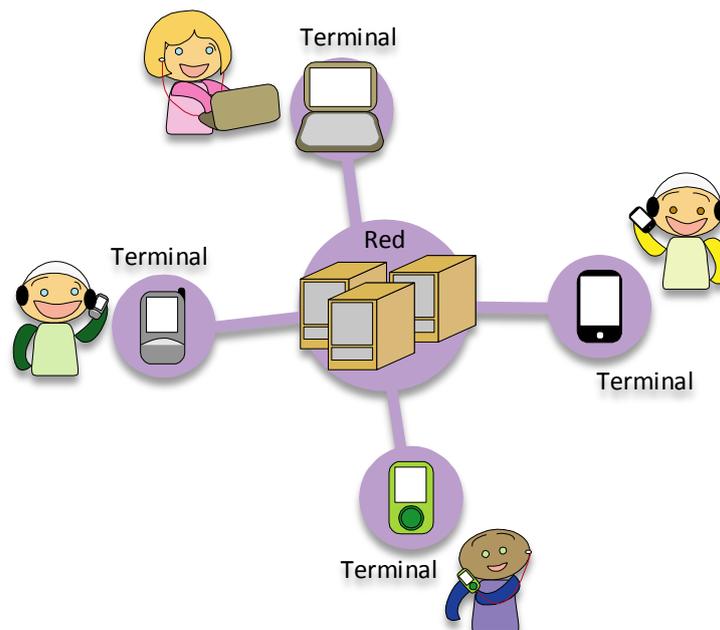


Figura 3. Arquitectura cliente-servidor para multiconferencia con diferentes terminales móviles.

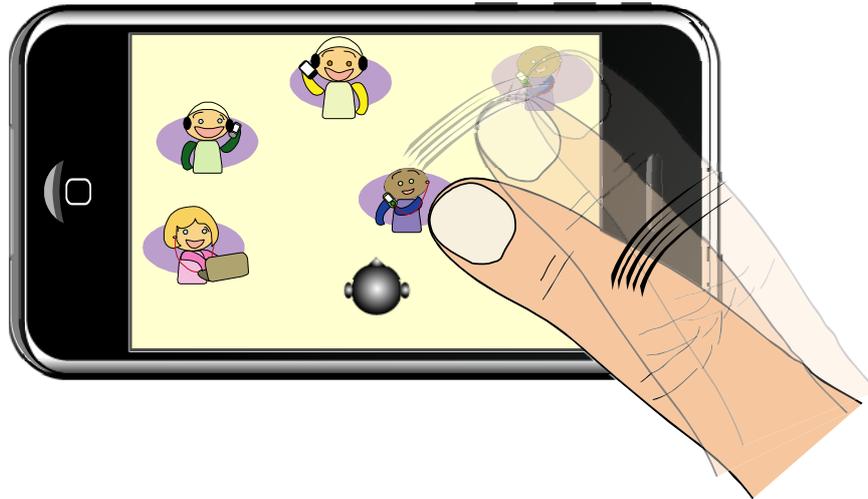


Figura 4. Concepto básico de la interfaz táctil de usuario para mover intuitivamente a los participantes por la sala.

1.5. Estructura de la memoria

La presente tesina se ha estructurado de la siguiente manera:

El capítulo 2 se trata de un estudio del estado de arte de la multiconferencia y los terminales móviles inteligentes. Se comienza con una breve historia de la evolución de la telefonía móvil desde sus inicios hasta hoy en día. Se continúa con el análisis de los nuevos teléfonos móviles inteligentes (*smartphones*), haciendo hincapié en los sistemas operativos móviles que utilizan e introduciendo un nuevo tipo de dispositivos derivados de los primeros, las *tablets*. Después se realiza un estudio de las actuales soluciones para multiconferencia por móvil que ofrece el mercado. Y se finaliza extrayendo las conclusiones que más van a afectar al tomar decisiones en este proyecto.

En el capítulo 3 se hace un estudio de los requisitos del proyecto. Es un fase esencial antes de comenzar cualquier desarrollo software. Se comienza describiendo la experiencia de usuario que se pretende obtener con este servicio. Después se analiza cómo se debe comportar el producto de cara a gestionar demostraciones de esta plataforma piloto. Por último se intenta enumerar todos los aspectos técnicos que nos pueden provocar limitaciones en el desarrollo para saber qué condiciones básicas debemos cumplir en todo momento.

El capítulo 4 describe la principales tecnologías que se han escogido para desarrollar este proyecto teniendo en cuenta los requisitos del capítulo anterior

que hay que cumplir. La primera de las tecnologías elegida es el códec de compresión Speex, que obtiene una gran tasa de compresión manteniendo una muy alta calidad de audio. La segunda es la técnica binaural de espacialización de sonido basada en HRTF, que simula el modo en que funciona el sistema auditivo humano. El último punto está dedicado al análisis más detallado de los dos sistemas operativos para móviles para los que se va implementar la aplicación: iOS (utilizado por los terminales iPhone y iPad) y Android (plataforma abierta a todos los fabricantes).

En el capítulo 5 se entra a fondo en la implementación final de las aplicaciones de este sistema de multiconferencia. El desarrollo comienza con el diseño de la interfaz gráfica para los terminales móviles, parte esencial para obtener un sistema con buena usabilidad. Se comentará la importancia de la homogenización para los dos sistemas operativos móviles, la estructura diseñada para obtener una interfaz intuitiva y la parte más artística pero no por ello menos importante de la interfaz. Después se explica cómo se han adaptado y optimizado los algoritmos y tecnologías escogidas, teniendo en cuenta las complicadas condiciones sobre las que deben funcionar. Por último se describe toda la parte de programación del proyecto que ha llevado a obtener tres aplicaciones finales: una aplicación servidor sobre PC y dos aplicaciones para terminales iOS y Android respectivamente, con las diferencias que conlleva la programación en cada una de estas plataformas.

El capítulo 6 es el último, y en él se analizan los resultados y trabajo futuro del proyecto. El primer punto resume los resultados obtenidos, explicando por qué ha sido un proyecto innovador, lo que ha aportado, lo costes en términos de trabajo y la calidad final conseguida. En segundo lugar se expone la repercusión que ha tenido este proyecto tanto en Telefónica, como en las ferias tecnológicas en que ha sido expuesto, como en las diversas apariciones que ha tenido en la prensa. El último punto del capítulo se dedica a analizar la evolución futura de este proyecto, tanto a corto plazo con mejoras del demostrador, como a largo plazo con las posibilidades que pueda haber de comercialización de un servicio así.

Capítulo 2

Estado del arte de la multiconferencia y los terminales móviles

2.1. Evolución de la telefonía móvil

La telefonía móvil es una de las tecnologías que ha revolucionado el mundo cambiando la forma de comunicarse del ser humano. Permite que cualquier persona esté conectada con el resto en casi cualquier lugar, dándole libertad de movimiento y ayudándole en distintas situaciones donde anteriormente hubiera estado incomunicado.

Se podría decir que la telefonía móvil nace de la combinación de dos tecnologías que muchísimo tiempo atrás también habían supuesto una revolución. Por una parte la invención del teléfono por Alexander Graham Bell en 1876 y por otra parte la invención de la radio (al fin y al cabo una comunicación inalámbrica) formalmente presentada por Guglielmo Marconi en 1894.

Si bien fue en la década de 1940 cuando aparecen los primeros radioteléfonos en EEUU, éstos eran sistemas muy arcaicos, por lo que se puede considerar que fue en la década de 1980 cuando nacen los primeros sistemas comerciales de telefonía móvil tal y como los entendemos hoy en día. Estos primeros sistemas, considerados la **primera generación (1G)** únicamente transmitían voz utilizando distintos sistemas de modulación analógica.

En la década de 1990 aparece la **segunda generación (2G)** de telefonía móvil caracterizada por la introducción de la digitalización de la voz para ser transmitida con nuevas modulaciones. Con la aparición de nuevos estándares como el GSM [5], la telefonía móvil se populariza rápidamente a nivel mundial. Con los nuevos protocolos de los sistemas 2G, además de transmitir mayor calidad de voz, se comienzan a transmitir también datos como fax y SMS, aunque siempre muy limitados en cantidad de información.

A comienzos del este nuevo siglo, en Europa y EEUU la tecnología 2G evoluciona a la **generación 2.5G**. Las novedades son las mayores velocidades de datos que ofrecen (hasta 114 Kbps con GPRS o 384 Kbps con EDGE) que permiten la aparición de servicios como los mensajes multimedia (MMS) que incorporan imágenes y sonidos en el mismo mensaje.

Poco más tarde irrumpe la **tercera generación (3G)** que no comienza a popularizarse hasta finales de la primera década del 2000. Básicamente nace con la intención de aumentar drásticamente el ancho de banda de datos (se puede llegar hasta los 7.2 Mbps) para permitir servicios como la conexión a Internet desde el móvil, la videoconferencia, la televisión y la descarga de archivos. La principal tecnología en que se basa la telefonía móvil 3G es el UMTS, nacido como sucesor del GSM para poder ofrecer servicios con mayor ancho de banda [6].

Esta es la tecnología que estamos disfrutando actualmente y que ha coincidido de forma paralela con la rápida evolución de los terminales móviles hacia lo que son denominados **teléfonos inteligentes** (o *smartphones*). Estos nuevos terminales se caracterizan por incorporar un sinnúmero de funciones más allá de las tradicionales llamadas telefónicas. Como mínimo permiten consultar el correo electrónico y tener un organizador personal, pero se suele poder instalar otros tipos de aplicaciones según el sistema operativo de que dispongan. A nivel de hardware, su potencia de cálculo no para de aumentar y disponen de teclados completos o pantallas táctiles como nuevas formas de interacción. Así se produce una perfecta simbiosis entre la telefonía 3G y los nuevos terminales inteligentes que sacan provecho de la conectividad que ofrece el 3G.

En los próximos años, se espera la implantación de sistemas de **cuarta generación (4G)** que volverán a incrementar drásticamente el ancho de banda de las comunicaciones por encima de los 100 Mbps. Tecnologías como WiMax y LTE proporcionarán acceso a nuevos servicios avanzados desde los terminales móviles.

Por otra parte, desde la aparición de las tecnologías 3G que dan acceso a la red de Internet desde el móvil, cobra interés el uso de **tecnologías de VoIP** (voz sobre IP) desde el móvil. Las técnicas de VoIP permiten transmitir la voz a través de Internet entre distintos terminales conectados a la red, para establecer así llamadas telefónicas sin usar los circuitos tradicionales de voz de los sistemas de telefonía fija o móvil. De esta forma, la voz, digitalizada, comprimida y dividida en paquetes de datos, se envía a través de Internet utilizando protocolos IP [7].

2.2. Terminales móviles inteligentes

2.2.1. Introducción

Durante la década de 1980 (desde el comienzo de la primera generación de telefonía móvil) y también la de 1990, los terminales móviles no sufrieron una gran evolución, si consideramos que durante casi 20 años los mayores cambios fueron una reducción progresiva de tamaño y la incorporación de algunas simples funciones extras como los SMS, despertador o agenda.

En los primeros años del 2000 los terminales comienzan a incorporar todo tipo de funciones cada vez más avanzadas, comenzando por la reproducción de música, cámara de fotos y acceso al correo electrónico. La potencia de sus procesadores va aumentando y van absorbiendo las funciones que hasta ese momento desempeñaban las PDAs (asistentes digitales personales).

No existe unanimidad en lo que es un terminal móvil inteligente (o *smartphone*) o las características que debe tener. Se podría considerar que la aparición del primer sistema operativo para móviles, Symbian OS en el año 2000, marca el inicio de este tipo de terminales que poco a poco fueron incorporando más aplicaciones así como el uso de otras tecnologías como el GPS o Wi-Fi.

Pero posiblemente la aparición del iPhone, el primer teléfono móvil de Apple en el año 2007, supuso el salto definitivo que popularizó ampliamente este tipo de teléfonos y aceleró la competencia de este nuevo modelo de negocio de terminales y aplicaciones. Se introdujo así la combinación de la primera gran pantalla táctil junto con un sistema operativo altamente intuitivo y una tienda con aplicaciones de terceros. En poco tiempo el mercado se llenó de diversas alternativas siguiendo esta combinación de características.



Figura 5. Teléfono móvil inteligente del fabricante HTC.

2.2.2. Sistemas operativos para móviles

Los actuales terminales inteligentes se suelen caracterizar por tener grandes pantallas (con alta resolución y táctiles), cámara, sistemas de posicionamiento GPS, conectividad Wi-Fi y 3G, gran potencia de cálculo y un sistema operativo muy visual y cómodo que permita la descarga de aplicaciones de terceros. Y es el sistema operativo (junto con sus tiendas de aplicaciones) lo que nos permite hacer la siguiente clasificación de los terminales actuales.

iOS

Es el sistema operativo creado por Apple para el iPhone en el 2007 (inicialmente llamado iPhone OS). En el año 2008 se lanza la tienda de aplicaciones App Store donde cualquier programador o empresa puede vender sus aplicaciones para iOS (después de que Apple apruebe la aplicación) [8]. Este sistema operativo no es abierto en el sentido de que solo es utilizado por los terminales de Apple, aunque tiene la ventaja de ser usado también en los dispositivos iPod Touch y en las *tablets* iPad de Apple. Una de las mayores quejas que ha sufrido es el hecho de no soportar la tecnología Adobe Flash.

Android

Es actualmente el mayor rival de iOS. Fue lanzado en el 2008 por Google como un sistema operativo abierto que puede ser utilizado por distintos fabricantes de terminales móviles. Dispone de la tienda de aplicaciones Android Market [9]. La ventaja de ser utilizado por muy distintos terminales a veces se convierte en la desventaja de no encajar perfectamente en algunos de ellos o que no todas las aplicaciones funciones igual en todos los terminales.

Windows Phone 7

Es el último sistema operativo de Microsoft para móviles que intenta entrar en este mercado. Propone una interfaz visual diferente dividiendo la pantalla en 7 áreas que se van activando dinámicamente con el uso. Tiene su propia tienda de aplicaciones llamada Windows Phone Marketplace [10]. Es todavía un sistema operativo muy reciente cuya evolución se desconoce.

Blackberry

Es el sistema operativo de RIM utilizado únicamente por sus teléfonos móviles, cuya primera versión salió en el año 2002. Tiene un software robusto, con funcionalidades profesionales, junto con el acceso a redes sociales tan demandado actualmente. Posee un sistema propio de mensajería instantánea que ha tenido

gran éxito y los terminales móviles se caracterizan por tener un teclado físico completo. Por supuesto también dispone de su propia tienda de aplicaciones llamada BlackBerry App World [11].

Symbian

La primera versión de este sistema operativo salió en el año 2000. Pero parece que no ha evolucionado tan rápido como iOS o Android y actualmente se ha quedado un atrás en la carrera de los terminales inteligentes. Está sustentado por el fabricante de móviles Nokia y dispone de la tienda de aplicaciones OVI [12].

Bada

Es el sistema operativo creado por Samsung para algunos de sus terminales móviles. Se lanzó en 2010 y es todavía muy reciente, habiendo pocos modelos que lo incorporen. Tiene una tienda de aplicaciones llamada Samsung Apps [13].

Cuota de mercado de los terminales móviles según su sistema operativo

Según la consultora Gartner [14], la venta de terminales móviles inteligentes aumenta significativamente año tras año (85 % durante el 2010 y 19 % en el primer trimestre del 2011 en comparación con el año anterior). Este incremento se está produciendo incluso en la actual época de recesión económica mundial donde otros productos tecnológicos disminuyen sus ventas.

La evolución actual de los teléfonos inteligentes muestra que Android se está expandiendo muy rápidamente, siendo el sistema operativo que más teléfonos vende actualmente. Según los resultados del primer trimestre del 2011 publicados por esta misma consultora [14], Android es actualmente líder con una cuota de mercado del 36 %. Si bien Symbian ocupa el segundo lugar con un 27 % de cuota, está retrocediendo muy rápido trimestre a trimestre. El tercer lugar es para iOS con un 17 % de cuota, aumentando progresivamente su presencia en el mercado. Con un 13 % de cuota, BlackBerry ocupa el cuarto lugar, aunque lentamente disminuyendo sus ventas. El resto de sistemas operativos se encuentran por debajo del 4 %.

Considerando que Symbian, además de estar retrocediendo a marchas forzadas, no cuenta con una tienda de aplicaciones madura y a la altura de sus competidores, podemos decir que los sistemas operativos más populares y con mejor mercado para sus aplicaciones son actualmente Android e iOS, seguido de BlackBerry. En un futuro se espera ver la evolución de Windows Phone 7, que todavía no es un competidor para los anteriores.

2.2.3. Tablets

Una *tablet* se podría definir como un ordenador personal portátil que no usa teclado ni ratón porque tiene una pantalla táctil para interactuar con él. Los primeros dispositivos de este tipo aparecieron de la mano de Microsoft a principios del 2000 con los llamados *Microsoft Tablet PC*. En un principio las *tablets* eran un derivado de los ordenadores personales y utilizaban los mismos sistemas operativos que los PC, no llegando a tener un gran éxito en esos primeros años.

Pero a partir del año 2010, potenciado por la aparición en el mercado del *tablet* iPad de Apple, cambia el concepto de estos dispositivos y comienzan a ganar gran popularidad entre los consumidores. Hoy en día una *tablet* no suele ejecutar un sistema operativo de PC, sino que utiliza los sistemas operativos para terminales móviles o derivados de estos. De esta forma consiguen ser un dispositivo mucho más fácil, robusto e intuitivo, pudiéndose considerar más como un terminal móvil inteligente de pantalla grande que como un ordenador de pantalla táctil. Al fin y al cabo son el producto de la convergencia entre los ordenadores y los móviles.



Figura 6. La *tablet* iPad 2 de Apple.

Las *tablets* tienen hoy en día diversos usos como puedan ser la lectura de libros, la ofimática, la navegación web, la reproducción de videos y música, los juegos, etc. Muchas de sus utilidades derivan de sus capacidades de conexión Wi-Fi e incluso 3G.

En cuanto a los sistemas operativos para *tablets*, en estos momentos se impone claramente iOS, sin gran competencia todavía, porque tiene con diferencia la mayor cuota de mercado (empujado por las grandes ventas del iPad) y también una mayor madurez (no es exclusivo para *tablets* sino compartido con los móviles iPhone). El resto de sistemas operativos para *tablets* como HoneyComb (variante de Android), WebOS o Meego todavía no se han consolidado, entre otros motivos, por ser muy recientes, y se espera que algunos de ellos no pueda sobrevivir dentro la fuerte competencia que se espera.

2.3. Soluciones comerciales de multiconferencia por móvil

Actualmente existen diversas soluciones comerciales para establecer una multiconferencia con terminales móviles, además de con telefonía fija como se ha hecho siempre. Suelen ofrecer la posibilidad de establecerla inmediatamente o reservando con antelación una sala virtual. Algunas soluciones funcionan solo por VoIP y otras permiten conectar también por GSM o telefonía fija. Las soluciones más simples suelen ser gratuitas mientras que el resto suelen facturar a través de un teléfono 900 o 902 y son ampliamente utilizadas a nivel empresarial.

A continuación vamos a ver las más importantes junto con sus características más destacables para nuestro interés.

Orientadas a empresas y autónomos:

Multiconferencia SIMULCOM de Movistar [15]

- Funciona con teléfonos fijos, RDSI y móviles.
- Permite audio o audio/video, desde 3 hasta 300 personas a la vez.
- Se puede iniciar inmediatamente o programando una reserva.
- Se puede facturar a cada participante o al coordinador.
- Se accede a través de un número 900 o 902.

PowWowNow [16], BTMeetMe [17], Arkadin [18]

- Realizan multiconferencias por teléfono con números 902 o por internet con funciones extra como la posibilidad de compartir documentos.
- Se pueden iniciar inmediatamente o programando una reserva.

Mirial ClearSea [19]

- Conecta con estándares H.323 y SIP haciendo video-multiconferencia.

- Funciona con Android, iOS y PC.
- Tiene un aspecto gráfico muy poco cuidado.

Orientadas al público general:

“Llamada a 3” clásica con terminales móviles

- Se inicia reteniendo una llamada para poder llamar a otro usuario y añadirlo.
- Habitualmente se tarifica por cada llamada nueva que hay que añadir.
- No es VoIP, si no llamada tradicional.

Skype [20]

- Solución VoIP que permite realizar audio-multiconferencia a través de un terminal móvil inteligente.
- Solo en la versión para PC se puede realizar también video-multiconferencia.

2.4. Conclusiones

Por una parte se ha visto como la evolución de los teléfonos móviles nos permite disponer hoy en día de **terminales muy potentes** para aplicar los procesados de señal que necesitamos, con una cómoda interacción para el usuario (**pantallas grandes y táctiles**) y la posibilidad de realizar nuestras propias aplicaciones para sus sistemas operativos. Así mismo, a partir de la aparición de las tecnologías 3G, tenemos **acceso a Internet desde el móvil**, pudiendo implementar nuestro propio flujo de datos VoIP para realizar nuevos servicios de voz.

En cuanto a sistemas operativos móviles, se ha visto como actualmente los mercados de aplicaciones más interesantes y con más futuro para programar y distribuir nuestras aplicaciones son los de **Android** y los de **iOS**, seguido de Blackberry.

Por otra parte, al analizar las **soluciones de multiconferencia** actuales en móviles, se puede llegar a varias conclusiones. En cuanto a la calidad de sonido, muchas tienen una **calidad pobre de voz** como la de la telefonía fija o la móvil GSM. Esto significa que muestrean la voz solo a 8 kHz. Solo unas pocas basadas en VoIP suben la calidad a 16 kHz. Lo que tienen en común todas es que **no realizan ninguna espacialización 3D del sonido**, es decir, el sonido es siempre monofónico. Por último, **no tienen una interfaz gráfica intuitiva** que permita identificar a los participantes que hablan en cada momento.

En referencia a la espacialización de sonido en llamadas con más de dos participantes, solo queda indicar que se sabe que el instituto Fraunhofer está desarrollando el estándar MPEG SAOC (Spatial Audio Object Coding). Este estándar está relacionado con la espacialización de sonido en conferencias permitiendo integrar varias señales de voz en una única trama para reducir el ancho de banda necesario en multiconferencias [21]. En cualquier caso, hasta el momento no se conoce que exista ningún producto o servicio que haga uso del MPEG SAOC. La solución que propone este estándar presenta además problemas de escalabilidad porque supone una sobrecarga importante de cálculo en el servidor.

Capítulo 3

Estudio de los requisitos del proyecto

3.1. Requisitos y plataformas

Como todo proyecto por encargo, es muy importante comenzar haciendo el mejor estudio posible de los requisitos antes de ponerse a desarrollarlo. Esto aclara mucho el camino a seguir y ahorra trabajo inútil durante el desarrollo.

En primer lugar fue necesario entablar diversas conversaciones con el contratante Telefónica para definir el funcionamiento del demostrador de multiconferencia inmersiva. Esto supuso la realización de una serie de propuestas que habíamos ideado y que ellos fueron corrigiendo según consideraban que eran sus necesidades. Básicamente estamos hablando de establecer la **experiencia de uso** que se quiere dar a los clientes del servicio, así como la forma en que los encargados de realizar demostraciones de este servicio van a **gestionar el sistema**.

En segundo lugar, y ya a modo más interno, se tuvo que realizar un análisis de las posibilidades técnicas de las que se disponía para lograr la experiencia de uso que se nos había pedido. Eso supuso detectar las **limitaciones tecnológicas** a las que nos enfrentábamos para poder decidir así que arquitectura, tecnologías y algoritmos nos convenía utilizar o desarrollar.

Plataformas para servidor y terminales móviles

Por último, aunque presente desde el principio del contrato, era importante saber sobre qué plataforma y tipo de terminales iba a funcionar el sistema, lo cual influye tanto en la metodología como en el volumen de trabajo. Con el contratante Telefónica se llegó al acuerdo de que el servidor que gestiona el sistema funcionase sobre cualquiera de las últimas versiones de Windows (XP, Vista y 7) y que los terminales fueran tanto de iOS (iPhone y iPad) como Android.

El hecho de que el funcione el **servidor sobre Windows** da la facilidad de montar una demostración inmediatamente con casi cualquier ordenador o portátil que se tenga a mano en un momento dado, reduciendo así el tiempo de gestión del encargado.

La elección de **iOS y Android** se debe a que son los dos sistemas operativos para móviles con mayor crecimiento actualmente (como se ha visto en el punto 2.2.2), cubriendo así una gran parte del mercado y aprovechando la popularidad de la que actualmente gozan. El hecho de añadir una *tablet* tenía un interés especial para demostrar como pueden interactuar estos nuevos dispositivos junto con los terminales móviles dentro del mismo sistema, además de poder evaluar también así las diferencias de interacción con una pantalla más grande y supuestamente cómoda. Se ha escogido la **tablet iPad de Apple** por ser actualmente la única consolidada ya que usa el mismo sistema operativo iOS y la tienda App Store que el iPhone (ver punto 2.2.3).

Si bien en una versión comercial de este sistema habría que considerar la extensión a otros sistemas más minoritarios como Blackberry o Windows Phone 7, para los objetivos de un demostrador piloto como éste no resulta rentable añadir más terminales móviles.

3.2. Experiencia de uso para el cliente del servicio

Con este demostrador, se pretende realizar una multiconferencia donde el cliente pueda colocar en un escenario, en tiempo real, al resto de los participantes. La facilidad de uso tiene que ser la clave, por eso se ha optado por aprovechar las capacidades táctiles de las pantallas de los terminales actuales. De este modo, intuitivamente el cliente solo tendrá que tocar y arrastrar con el dedo a cada participante al sitio del escenario donde desee escucharlo.



Figura 7. Interfaz táctil para mover a los participantes.

La voz de los participantes se escuchará a través de los auriculares (con técnicas más avanzadas que estéreo tradicional) en función de su posición en el escenario, permitiendo al cliente discernir con más facilidad quién está hablando en cada momento ya que les escuchará como si se encontrara con ellos en ese momento en la misma sala. Así mismo, para aumentar el realismo, se debe mejorar la calidad de voz respecto a la de las llamadas telefónicas tradicionales.

Como se trata de un demostrador que tiene que resultar muy ágil allá donde se pruebe, la curva de aprendizaje debe ser extremadamente rápida porque los potenciales clientes van a disponer de muy pocos minutos para aprender a usarlo. Por eso aunque en una versión comercial deberían existir un mayor número de ajustes y posibilidades, en este caso se ha tenido que simplificar al máximo el modo de acceso y uso para que resulte absolutamente intuitivo.

Para entrar en la sala de la multiconferencia los usuarios simplemente tendrán que arrancar la aplicación donde se les pedirá usuario y contraseña. Automáticamente, el recién incorporado entrará en la sala donde verá al resto de participantes con la foto o avatar que hayan elegido para representarles. Esta representación gráfica resulta de gran ayuda para poder identificar inmediatamente a quién habla en cada momento, aún cuando no se conozcan previamente sus caras y voces.

Por último, se va a permitir que el usuario personalice la escena donde quiere que transcurra la multiconferencia, eligiendo el escenario que más le guste entre un abanico de ambientes de empresa, familiares o de amigos.

3.3. Gestión del demostrador

Detrás de cada demostración de este sistema piloto que se realice habrá algún encargado de gestionarla. Aunque este encargado pueda tener conocimientos técnicos y disponga de un manual de uso, se le debe facilitar al máximo la tarea de gestión.

Las condiciones en que se realizan las demostraciones son a menudo hostiles, por ejemplo cuando se presentan en ferias tecnológicas o a visitas inesperadas. En ocasiones se dispone de muy poco tiempo de preparación, puede asignarse de repente un encargado nuevo, el equipamiento puede cambiar, la conectividad en un sitio puede ser limitada, etc.

Todo esto nos obliga a que el sistema de gestión sea lo más robusto posible, así como sencillo y portable. Se establece que para que funcione todo el sistema se necesite únicamente un ordenador o portátil como servidor de las comunicaciones,

una conexión a Internet para éste y una aplicación que se le instale y ejecute sin más.

No se debe obligar a que el PC tenga en principio ninguna característica especial, se tiene que poder utilizar cualquiera de los que se venden comúnmente desde hace un par de años que funcionen con Windows XP, Vista o 7. Este servidor no tiene por qué estar situado cerca de los clientes, sino que puede encontrarse en cualquier otra ciudad al estar conectado a Internet.

La aplicación servidor debe funcionar únicamente “haciendo click” en un botón de encendido, sin que el encargado necesite habitualmente ajustar ningún parámetro más, para evitar así errores accidentales. Es importante que en el servidor vaya saliendo por pantalla la información de lo que está sucediendo (usuarios que conectan, cuándo, desde donde, etc.) para que el encargado pueda anticiparse a cualquier incidente no previsto en la demostración. Así mismo, si estuviera ocurriendo cualquier situación anómala, la aplicación servidor deberá informar en la medida de lo posible del error y su posible causa para ayudar al encargado a solucionarlo cuanto antes.

3.4. Aspectos técnicos y limitaciones

La implementación del proyecto propuesto sobre un terminal móvil presenta una serie de retos y dificultades. Se debe mantener una conversación con muchos participantes a la vez y mejorar la inmersividad con algoritmos de audio espacial, todo desde dispositivos móviles. Como veremos, las limitaciones técnicas se van a ir encadenando.

Conectividad de los terminales y ancho de banda

En la actualidad las llamadas telefónicas tanto en 2G (GSM) como en 3G (UMTS) utilizan el concepto de conmutación de circuitos. Esta tecnología dispone de unos protocolos estándar que en principio no están preparados para añadir servicios más complejos de voz como el presentado en este proyecto. Así mismo los fabricantes de terminales móviles no proporcionan las herramientas de desarrollo necesarias para poder modificar esta capa de protocolos relacionada con el propio sistema de telefonía.

Por lo tanto, ante la dificultad de desarrollar el prototipo empleando la parte de conmutación de circuitos de un terminal móvil, es necesario utilizar las capacidades de **transmisión de datos en redes IP** de los terminales actuales mediante **red móvil 3G** para establecer el canal de comunicación de la

multiconferencia. Por ello se han de usar técnicas que permiten transmitir la voz sobre redes IP (llamadas genéricamente VoIP).

El propósito final y más exigente es que el usuario tenga la máxima libertad de movimiento, pudiendo utilizar el servicio desde cualquier sitio con cobertura 3G. Pero por supuesto, si se implementa la transmisión de la voz por redes IP, el cliente también podrá acceder al servicio a través de una **conexión Wi-Fi** a Internet.

Pero volviendo a la condición más exigente de que el terminal acceda con una conexión 3G, esto nos obliga a **minimizar el ancho de banda** requerido para la comunicación. Esto es lógico, no solo por el coste económico que supone el flujo de datos en una conexión móvil, sino también para mantener máxima la calidad de audio aún en las condiciones de baja cobertura de 3G que reducen las velocidades de transmisión.

Capacidades del servidor

A diferencia de las llamadas telefónicas uno a uno, en una multiconferencia se requiere que cada terminal reciba la señal de varios participantes. Si queremos utilizar poco ancho de banda, sobre todo en el canal de subida de los terminales (que es el más limitado), tenemos que descartar las conexiones *peer-to-peer* entre los móviles porque obligaría a que cada uno de ellos enviara varios flujos de voz idénticos al resto de participantes, saturando el canal de subida.

Para hacer un uso mucho más eficaz del ancho de banda, se hace necesaria la participación de un servidor de forma que recoja la señal de voz de cada terminal y se encargue de enviársela al resto de participantes. Esto logra que cada terminal solo tenga que enviar un flujo de señal de voz al servidor, quien se encargará de distribuirlo. Como se verá en el siguiente capítulo, se añadirán algunas técnicas más para reducir el volumen de datos.

Las gestiones que realice el servidor con los flujos de audio deben ser **optimizadas al máximo para que utilicen muy poca CPU**. Esto es importante a corto plazo para hacer funcionar el servidor del demostrador con muchos usuarios sobre un PC común de bajo precio. Pero también nos interesa a largo plazo que las técnicas que utilicemos sean de bajo coste computacional para permitir su escalabilidad a miles de clientes en un posible comercialización futura.

Limitaciones de los terminales móviles

Los terminales actuales disponen de una gran potencia de cálculo. Cuando se diseñen las técnicas y algoritmos de tratamiento de la señal, deben poder ser

ejecutadas en tiempo real al menos por los móviles inteligentes de gama media actuales.

Pero no basta con que los cálculos no bloqueen el dispositivo, el uso de la CPU se debe minimizar todo lo posible para disminuir así el **consumo de la batería**. Una aplicación que agota en poco tiempo la batería no da un buen servicio al usuario.

El consumo de la batería depende de diversos factores. Algunos como la pantalla y la salida de audio en este caso son inevitables, en cambio en otros como el **uso de CPU** y la **cantidad de datos que se transmiten** y reciben sí que tenemos margen de maniobra. Así que ésta es otra razón más para intentar reducir el flujo de datos necesario.

Calidad de la conversación

Se ha explicado la importancia de realizar un procesado de la señal con muy bajo coste computacional y que utilice flujos de datos ligeros. Pero no podemos olvidar el propósito original de este proyecto que se trata de mejorar la calidad de las conversaciones de una multiconferencia, así que hay varios puntos a los que no se puede renunciar:

- Por una parte no podemos utilizar señales de voz de 8 kHz como en las llamadas tradicionales, es imprescindible utilizar al menos señales de 16 kHz para que la voz no suene “enlatada” y parezca más real.
- Obviamente se debe realizar una compresión fuerte de la señal de audio antes de transmitirla pero se debe buscar un codificador específico para voz que a su vez no estropee la calidad de audio que se va a percibir.
- Dado que se van a llegar a escuchar a varios participantes a la vez, el ruido ambiente que se acumula de cada uno se convierte en un problema importante y hace imprescindible el uso de técnicas de reducción de ruido.
- Por último, sin consumir mucho procesado, los algoritmos de espacialización de las diferentes señales de voz tienen que ser realmente buenos para que el usuario tenga una sensación de presencia real en la sala y pueda aprovechar sus capacidades naturales de focalización direccional de la escucha.

Capítulo 4

Tecnología utilizada

4.1. Introducción

Tras el análisis de los requisitos y limitaciones técnicas de proyecto, estamos en condiciones de explicar el funcionamiento de las principales tecnologías elegidas para su implementación así como las características de los sistemas operativos móviles que hemos de usar.

Procesado digital de audio

A nivel de procesado de la señal, el proyecto tiene dos claros pilares tecnológicos: el uso de **codificadores de voz de alta calidad**, que permiten una transmisión óptima de los flujos de voz a través de IP, así como el uso de **algoritmos de sonido 3D**, que permiten aplicar técnicas de audio espacial al sonido recibido.

- En relación al codificador, se ha utilizado el **códec Speex**, uno de los más avanzados que además tiene la particularidad de ser de código abierto, estando libre de patentes y permitiendo mejoras y adaptaciones del mismo a la aplicación final en la que va a ser empleado. Cabe indicar que en el afán de obtener una buena calidad de audio, se complementado el procesado de audio aplicando un **reductor de ruido ambiente** previo al codificador. Esto no solo aumenta la inteligibilidad, al haber sonido de tantos participantes a la vez, sino que también mejora las prestaciones del compresor de voz.
- En relación a las tecnologías de **sonido espacial**, se ha empleado **tecnología binaural (HRTF)** para conseguir la sensación de sonido espacial, que es la técnica más realista actualmente para la reproducción de audio en 3D sobre auriculares. Las señales de voz que llegan de los participantes de la multiconferencia son procesadas junto con la información de su posición en la escena, para de esta manera, mediante técnicas de procesado digital, producir

un sonido espacial muy realista que es reproducido mediante los auriculares estándar que incluyen los teléfonos móviles inteligentes.

Protocolos de transmisión

La transmisión de las señales ya se ha explicado que se iba a realizar sobre el protocolo IP de Internet por la imposibilidad de modificar los protocolos de las llamadas telefónicas. Hay que distinguir dos tipos de señales:

- Por una parte las **señales de voz**, que básicamente son tramas de audio comprimidas que se envían entre terminales y servidor. Para estas tramas se ha decidido utilizar el **protocolo UDP** porque permite trabajar a una gran velocidad sin tener que establecer una conexión punto a punto. Con UDP no está garantizada la recepción en destino de todos los paquetes, pero gracias al códec Speex esto no es un problema porque es robusto frente a pérdidas de paquetes.
- Por otro lado tenemos los **mensajes de gestión** de la comunicación que se tienen que intercambiar entre terminales y servidor. En este caso se ha optado por transmitir este tipo de mensajes sobre el **protocolo TCP**. Estos mensajes de señalización, aunque ocupan muy poco, son esenciales para establecer las comunicaciones, por lo que resulta imprescindible asegurar al máximo su recepción usando TCP.

4.2. Códec de compresión Speex

La función principal y para la que fue creada Speex es la **compresión de voz**, para la cual no existía **código libre** adecuado. A diferencia de otros codecs, Speex no se ha diseñado para trabajar con los protocolos de conmutación de circuitos de la telefonía fija o la telefonía móvil clásica, sino para redes de paquetes y aplicaciones de **voz sobre IP (VoIP)**. Por supuesto también sirve para compresión de voz en archivos [22][23][24].

El códec Speex ha sido diseñado para ser muy flexible y permitir un gran margen de calidad de voz y *bit-rate*, basándose en el algoritmo CELP. Esto significa que Speex no solo es capaz de comprimir **voz de bajo ancho de banda** (la calidad telefónica de **8 kHz** de muestreo) si no también de **gran ancho de banda** a frecuencias de muestreo de **16 y 32 kHz**.

Al haber sido diseñado para VoIP en lugar de para protocolos como GSM significa que Speex es **robusto frente a la pérdida de paquetes**, pero no frente a paquetes corruptos. Esto se basa en la suposición de que en VoIP, los paquetes o bien llegan sin alteración o bien no llegan. Como Speex está dirigido a un gran

abánico de dispositivos, tiene una baja complejidad (aunque ajustable) y utiliza poca memoria.

Podemos resumir las siguientes características como las más interesantes de la compresión Speex:

- Es un software abierto y libre de patentes.
- Se integra la voz de banda estrecha (8 kHz) y ancha (16 kHz) usando un flujo de bit embebido.
- Admite frecuencia de muestreo ultra-ancha a 32 kHz.
- Gran margen de *bit-rate* desde 2.15 kbps hasta 44 kbps.
- Cambio de *bit-rate* dinámico (AMR) y posibilidad de *bit-rate* variable (VBR).
- Detección de actividad de voz (VAD) y transmisión discontinua (DTX).
- Complejidad variable.

4.3. Sistema binaural HRTF de espacialización de sonido

Los sistemas tradicionales de audio estéreo son capaces de situar un sonido en un punto entre dos altavoces o auriculares. En estos casos, mediante una sencilla operación de *panning* se le da a cada sonido una ganancia distinta en cada altavoz y se consigue la sensación aproximada de que las fuentes se encuentran más a la derecha o a la izquierda.

Pero existen sistemas y algoritmos de sonido real 3D que son capaces de colocar los sonidos en cualquier posición del espacio (derecha/izquierda, arriba/abajo, cerca/lejos) utilizando únicamente dos auriculares, a costa de realizar unas operaciones más complejas que en un simple estéreo [25]. Se trata de sistemas binaurales de espacialización de sonido que intentan simular de forma virtual el funcionamiento del oído humano en la vida real y cómo éste es capaz de discernir la dirección de las fuentes a partir de las diferencias relativas en el sonido que recibe en cada oído.

Psicoacústica del sonido espacial

Ya hace unos 100 años, Lord Rayleigh desarrolló una teoría para explicar cómo el sistema auditivo humano escucha en 3D. Las dos razones principales y que más influyen en la **percepción del audio espacial en el plano horizontal** son [26]:

- La “**Diferencia de Tiempo Interaural**” (ITD). Considerando que el sonido avanza a una velocidad aproximada de 343 m/s, una onda de sonido proveniente desde una determinada dirección llegará antes a un oído que al otro por la distancia extra que tendrá que recorrer (salvo que el sonido esté centrado justo enfrente). Conociendo la separación entre los oídos y despejando una sencilla ecuación, podríamos prever la diferencia en tiempos de llegada en cada oído que se producirá según la dirección de la fuente.
- La “**Diferencia de Nivel Interaural**” (ILD). Cuando las ondas de sonido llegan a los oídos, sufren una difracción por la cabeza del oyente, que obstruye el camino hacia el oído que se encuentra al otro lado de la cabeza. Este obstáculo para el sonido produce una importante diferencia de nivel entre las señales de los dos oídos. Esta diferencia de nivel es altamente dependiente de la frecuencia. A bajas frecuencias la longitud de onda del sonido es larga respecto al diámetro de la cabeza, por lo que casi no hay diferencia en el nivel de la señal recibida en cada oído. En cambio, a altas frecuencias, como la longitud de onda es corta, se pueden producir 20 dB o más de diferencia de presión en cada oído.

La teoría enunciada por Rayleigh asegura que **la diferencia de tiempo y nivel interaural son complementarias**. A bajas frecuencias (por debajo de 1.5 kHz) hay poca diferencia de nivel pero la diferencia de tiempo desplaza la forma de onda solo una fracción de ciclo, lo cual el oído humano puede detectar fácilmente. A altas frecuencias (por encima de 1.5 kHz), se produce el desplazamiento de varios ciclos en la forma de onda, por lo que la diferencia de tiempos no ayuda a distinguir la dirección, pero la gran diferencia de niveles entre oídos sí que ayuda a distinguirla.

Se ha visto el modo en que el oído humano distingue la dirección de llegada del sonido en el plano horizontal. Pero para distinguir la **dirección de llegada en el plano vertical**, en altura, las claves que lo producen son mucho más sutiles. El principal factor que nos hace distinguir la elevación de una fuente es la resonancia tan particular que produce nuestro pabellón auditivo, que amplifica ciertas frecuencias en detrimento de otras según la altura de la fuente. Esto significa que podríamos simular la elevación de una fuente aplicando un filtrado frecuencial adecuado a las señales (algo muy dependiente de cada individuo). En cualquier caso hay que destacar que la capacidad del oído humano para distinguir la altura

de una fuente es muy limitada e inferior a la capacidad para hacerlo en el plano horizontal.

Por último, las capacidades del ser humano de distinguir la **distancia a que se encuentran las fuentes** de sonido se deben a múltiples y complejos factores entre los que destacan: diferencias de nivel de la señal al acercarse, diferencias del ángulo percibido cuando el oyente desplaza la cabeza y la relación de ganancia entre el sonido directo y el sonido reverberado por la sala.

Sistemas basados en HRTF

Los métodos de espacialización sonora basado en HRTF aprovechan los fenómenos psicoacústicos explicados anteriormente a través de la **respuesta al impulso de cada oído según la posición de la fuente acústica** [25]. HRTF es el acrónimo de “Función de Transferencia Relativa a la Cabeza” que se puede obtener bien **mediante mediciones reales o modelándolo** a partir de ciertos parámetros físicos de los individuos. Un HRTF resume como se ve afectada la señal en amplitud y fase para cada frecuencia en cada oído según la dirección de llegada de la fuente.

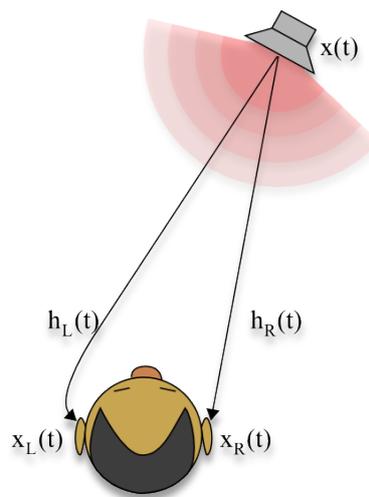


Figura 8. Diferente respuesta al impulso según el camino entre la fuente y cada oído.

Una vez se conoce el HRTF del oído izquierdo y derecho, en principio se podría sintetizar las señales binaurales a partir de fuentes de audio monoaurales. Uno de los problemas es que **el HRTF es dependiente de cada individuo**, de sus características físicas, por lo que un sistema perfecto debería ser distinto para cada individuo basándose en mediciones personalizadas. En la práctica esto no es posible y se suele buscar una solución de compromiso utilizando HRTF genéricos que den una buena sensación a la mayoría de usuarios.

Según estas técnicas, se puede realizar un **procesado binaural a una señal monofónica** para obtener las señales izquierda y derecha de cada oído que simulen la procedencia espacial de la fuente. Una de las formas de realizar este procesado es aplicando **algoritmos que calculen los retardos en tiempo** necesarios para cada oído y que **filtren en frecuencia la señal** para darle la ganancia adecuada según la posición virtual de la fuente. La complejidad de estos algoritmos se puede ajustar según el grado de realismo que se pretenda conseguir y los grados de libertad de la fuente que se quieran simular, teniendo en cuenta, por ejemplo, que nuestra capacidad para localizar la elevación es muy limitada.

4.4. Sistemas operativos para móviles

4.4.1. iOS

Originalmente llamado *iPhone OS*, iOS es el sistema operativo para dispositivos móviles de Apple. Actualmente es utilizado tanto por los teléfonos iPhone, como por los dispositivos iPod Touch y las *tablets* iPad [8].

Este sistema operativo fue lanzado por Apple en el año 2007 junto con su primer teléfono móvil iPhone. Cuando posteriormente aparecieron los dispositivos iPod Touch (también en 2007) y iPad (en 2010) también compartieron el mismo sistema operativo. Desde su nacimiento, sufre una actualización mayor aproximadamente una vez por año.

De cara al usuario, iOS dispone una **interfaz cómoda e intuitiva** basando su interacción en gestos multitáctiles. Los principales elementos de su interfaz son deslizadores, interruptores y botones que proporcionan una respuesta fluida. En la pantalla principal se sitúan los iconos de las aplicaciones que puede ser extendida lateralmente por más pantallas para ubicar más aplicaciones. En la parte inferior de la pantalla se sitúa el *dock* donde se “anclan” las aplicaciones que se usan más frecuentemente. En la parte superior de la pantalla se sitúa la barra de estado donde se muestran datos como la hora, la batería o la intensidad de la señal.

De cara al programador existe un **kit de desarrollo de software** (SDK) desde el año 2008 que permite a terceros desarrollar aplicaciones para iOS y publicarlas en la tienda de aplicaciones App Store. El SDK incluye numerosas utilidades como simulador de iPhone para Mac u otras herramientas de depuración para el dispositivo. El entorno principal de desarrollo, al igual que para Mac OS X, es el XCode, con muchos años en el mercado y un gran estado de madurez [27].

El lenguaje de programación para desarrollar aplicaciones es **Objective-C**, una variante especial de C orientado a objetos. Así mismo es posible introducir código en C nativo. Las aplicaciones deben ser compiladas para la arquitectura ARM, que es la que utilizan actualmente los dispositivos móviles de Apple.

Para comenzar a desarrollar aplicaciones para iOS primero es necesario ser aceptado en el *iOS Developer Program* pagando una cuota anual. Esto implica a su vez la aceptación de numerosos términos y condiciones incluyendo importantes cláusulas de confidencialidad. Los desarrolladores pueden distribuir sus aplicaciones bien en una modalidad interna y cerrada para su empresa o bien publicarlas en la **tienda App Store** para que cualquier usuario las descargue. En este último caso, el desarrollador decide si las ofrecen gratuitamente o el precio que desea que tengan (a partir de 0,79€) de los cuales recibirá el 70 % de las ganancias [28].



Figura 9. Pantalla principal del sistema operativo iOS.

4.4.2. Android

Android es un **sistema operativo abierto** para dispositivos móviles, creado por Google, que es utilizado por multitud de fabricantes de terminales móviles. Se ha popularizado fundamentalmente entre los teléfonos inteligentes, pero también puede ser utilizado por *tablets*, reproductores MP3, netbooks y otras clases de dispositivos. Es un sistema operativo basado en GNU/Linux cuyo código liberó Google bajo la licencia Apache, que es libre y de código abierto [9].

El sistema Android fue lanzado en el año 2008 como respuesta al sistema operativo iOS de Apple, pero con la diferencia de ser un sistema abierto al que puede tener acceso cualquier fabricante de terminales. Las actualizaciones son lanzadas a un ritmo rápido por parte de Google aunque de momento no todos los fabricantes las soportan a dicho ritmo, lo cual ha sido numerosas veces criticado.

Actualmente es el sistema operativo **con mayor presencia** entre los teléfonos inteligentes a nivel mundial [14].

Para el usuario, la **interfaz gráfica** ha sido diseñada para ser muy intuitiva y basada sobre todo en el uso de las pantallas táctiles. Aunque la interfaz mantiene siempre unas partes básicas comunes, no es igual en todos los dispositivos, no solo por sus distintas capacidades de cálculo, tamaño de pantalla, etc., si no también porque cada fabricante puede personalizarla a su gusto.



Figura 10. Pantalla principal de Android en un terminal HTC.

Para el programador, existe un **kit de desarrollo de software** (SDK) que Google ofrece gratuitamente para su descarga y que cuenta con varias herramientas útiles para la depuración de las aplicaciones. El lenguaje principal para desarrollar las aplicaciones es **Java**, si bien es posible insertar código C nativo en ciertas capas inferiores [29].

A diferencia de iOS y otros sistemas para móviles, Android se desarrolla de forma abierta, pudiéndose acceder tanto al código fuente como al listado de incidencias para ver los problemas no resueltos o reportar problemas nuevos. Ahora bien, para que Android funcione sobre un determinado terminal son necesarios ciertos controladores para el hardware que pertenecen a los fabricantes y no son públicos. Esta es una de las razones por las que no se puede instalar una nueva versión de Android en un dispositivo hasta que el fabricante no adapta los nuevos controladores.

El “ecosistema” Android dispone de la tienda oficial de aplicaciones **Android Market** [30] (y existen otras no oficiales), donde un desarrollador puede publicar sus aplicaciones pagando una cuota de US\$25. Las aplicaciones se puede ofrecer gratis o poner un precio, del cual el autor recibirá un 70 % de las ganancias.

Capítulo 5

Desarrollo de las aplicaciones del sistema

5.1. Introducción

Hasta el momento se han expuesto los requisitos que tiene que cumplir el sistema a desarrollar así como las tecnologías que se ha decidido utilizar. Pero el grueso del proyecto es el desarrollo del software final.

La interfaz gráfica de la aplicación en los terminales móviles es una de las partes importantes del proyecto porque se va a valorar que el producto obtenido sea altamente intuitivo, cómodo y agradable para el usuario. Esto supone realizar un **diseño bien estudiado de la interfaz** antes de comenzar la programación.

Otro aspecto muy importante antes de comenzar a programar líneas de código es tener muy claro cómo se van a **optimizar y adaptar las técnicas claves** para este sistema de multiconferencia. Estamos hablando de los ajustes del códec Speex óptimos para nuestro caso, el diseño de las tramas de audio y de gestión que se van a transmitir y la optimización de los algoritmos binaurales HRTF para la espacialización del audio. Para tener éxito en este proyecto, hay que conseguir mantener lo más alta posible la calidad y realismo del audio mientras se intenta minimizar el uso de CPU y el ancho de banda. Pero también es cierto que algunos de los ajustes se tendrán que finalizar haciendo distintas pruebas de código mientras se va desarrollando.

Desde el punto de vista de la **programación del software** tendremos que desarrollar **tres aplicaciones finales**:

- La primera es la aplicación servidor que se va a ejecutar sobre un PC y que va a gobernar toda la multiconferencia.

- Las otras dos aplicaciones son las que se van a ejecutar en los terminales móviles iOS (para iPhone y iPad) y Android respectivamente. Si bien estas dos aplicaciones tienen la misma funcionalidad, a nivel de programación casi no pueden reaprovechar código, porque la primera se debe programar en Objective-C y la segunda en Java, sumado a las correspondientes limitaciones y diferencias que impone cada sistema operativo.

5.2. Diseño de la interfaz gráfica

5.2.1. Homogeneidad entre los sistemas operativos

Como se ha comentado anteriormente, el sistema de multiconferencia se va a implementar para iOS (iPhone y iPad) y Android. Aunque se trata de dos sistemas operativos para dispositivos móviles inteligentes que por la tanto tienen puntos en común en su filosofía de uso, también tienen muchas diferencias.

Cada sistema operativo ofrece al desarrollador de aplicaciones, a través de su SDK, una serie de facilidades para crear elementos gráficos típicos como botones, deslizadores, menús, siendo el aspecto y filosofía de estos algo característico de cada sistema.

Si bien queremos respetar algunos detalles de cada sistema a los que sus usuarios están acostumbrados, también pretendemos mantener una imagen de homogeneidad de la aplicación a través de los distintos sistemas operativos, de momento para iOS y Android, pero que podrían ser más en un futuro. Esto **afianza la imagen de marca y la filosofía de la aplicación**, a la vez que para el usuario **facilita la migración** de un sistema a otro reduciendo el aprendizaje necesario. Así que buscamos ese punto de equilibrio donde las aplicaciones parecen iguales y se utilizan igual en todos los terminales pero a su vez respetan algunos detalles característicos ya sea de iOS o Android.

Este grado de homogeneidad lo conseguimos por una parte estableciendo una **secuencia de uso y estructura de pantallas común** y por otra creando una serie de **elementos artísticos comunes** para cada interfaz.

En cuanto a los elementos que van a quedar distintos, los limitamos básicamente a los menús secundarios, que aunque dispondrán de las mismas opciones en ambas plataformas, se va a permitir que su aspecto gráfico sea el que posee por defecto cada una y al que están acostumbrados los usuarios.

5.2.2. Estructura de la interfaz gráfica

La estructura de la interfaz gráfica va a marcar la secuencia de uso de la aplicación y por lo tanto su usabilidad. Como se ha repetido en varias ocasiones, queremos que sea fácil e intuitiva como tienen que ser las aplicaciones para móviles que tienen éxito. Pero su simpleza es importante, más si cabe, si tenemos en cuenta que se trata de un demostrador que va a probar mucha gente diferente disponiendo de muy poco tiempo.

La **pantalla principal** se ha diseñado de forma muy visual para que se entienda a la primera. Va a aparecer una imagen de fondo representando una sala donde van a estar los participantes. Sobre esta sala estarán los avatares de cada participante, con un tamaño grande para que el usuario pueda tocarlos y arrastrarlos a la posición que desee escucharlos. Esta pantalla se completa con tres botones muy claros: uno grande para salir de la conversación, uno pequeño para entrar en un menú de ajustes y otro pequeño para silenciar nuestro micrófono.

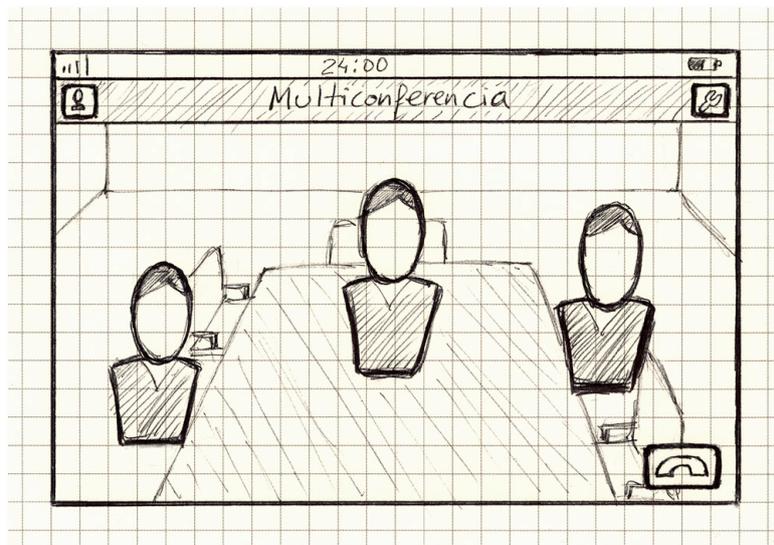


Figura 11. Boceto de la pantalla principal.

En la estructura de pantallas desde que se inicia la aplicación, solo se ha querido poner un **pantalla inicial** previa a la principal, y es en la que pide el nombre de usuario que va a entrar en la sala. Esta pantalla es imprescindible para que cada usuario entre con un avatar y nombre que lo identifique ante el resto. La pantalla se completa con una petición de contraseña opcional y un botón para ir a un menú de ajustes opcionales. Si bien en una versión comercial del sistema el usuario debería elegir a quién llamar o en que sala entrar, para dar agilidad a las demostraciones en cuanto el usuario pone su nombre entra a una sala única donde están el resto de participantes.

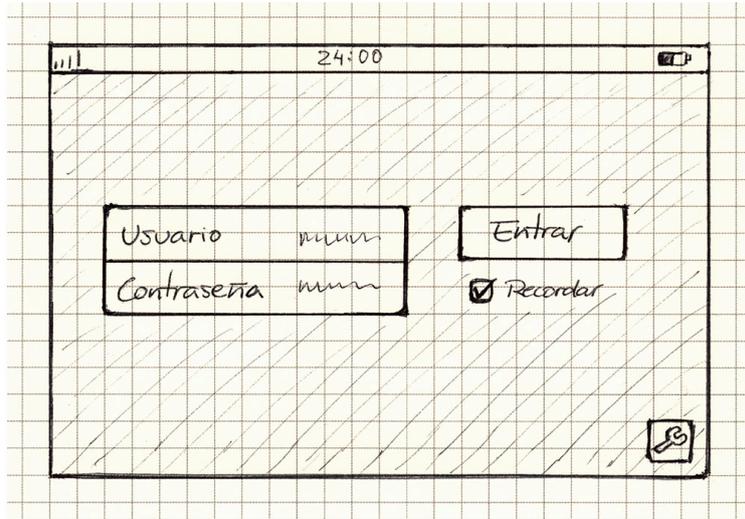


Figura 12. Boceto de la pantalla inicial.

La estructura se completa con dos **pantallas de ajustes** distintas y que habitualmente no son necesarias usar:

- A la primera se accede desde la pantalla de petición de nombre para entrar y normalmente no la utilizará el usuario sino el encargado de preparación de la demostración. En este menú se puede cambiar el idioma de la aplicación (Español/Inglés), elegir el servidor al que conectar (pueden haber varios activos por razones logísticas o de robustez) y activar una mejora de la transmisión para redes 3G (no necesaria si se usa sobre Wi-Fi).
- A la segunda se accede desde la pantalla principal de la sala virtual y este menú sí que está dirigido al cliente. Básicamente le permite ver su nombre y foto, elegir si quiere ver los avatares de los participantes como fotos o como dibujos y seleccionar la imagen de la sala que le guste más. Se ofrecen tres tipos de salas (corporativa, familiar y de amigos) en estilo foto o dibujo.

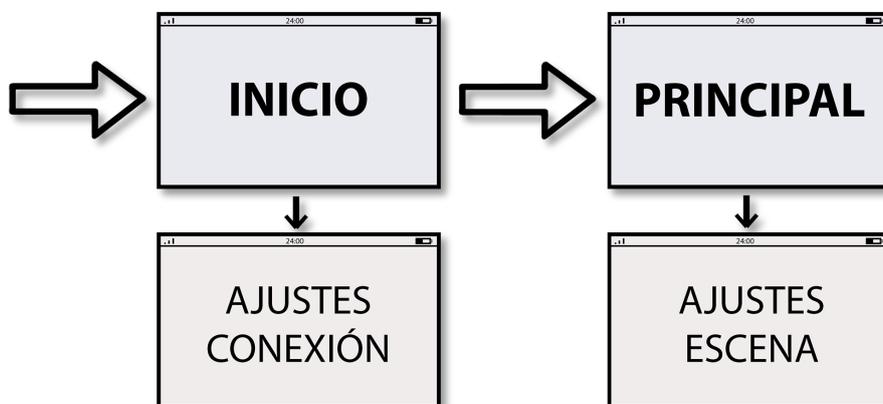


Figura 13. Estructura de pantallas de la aplicación para el móvil.

5.2.3. Arte de la interfaz gráfica

Se ha dedicado un gran esfuerzo en que los elementos de la interfaz tengan una alta calidad gráfica. Al fin y al cabo son la **carta de presentación de la aplicación** cara a los usuarios y se sabe que les influye en la experiencia final. Además, al realizar un trabajo para una empresa como Telefónica, es necesario entregarles una aplicación de aspecto impecable porque no pueden descuidar nunca la imagen que van a ofrecer a los usuarios en cualquier situación.

Casi todos los elementos gráficos han tenido que ser **creados expresamente para esta aplicación** por dos motivos. En primer lugar no se podían usar muchos de los elementos propios ya existentes de cada plataforma porque se pretende **mantener la homogeneidad** entre ellas. En segundo lugar porque la aplicación ha de mantener una **imagen de marca** que la identifique a simple vista con Telefónica o más concretamente con los productos Movistar.

Como toda gran empresa, Telefónica realiza un gran esfuerzo en mantener su imagen de marca (en este caso Movistar) hasta el más mínimo detalle en todos los aspectos gráficos de los productos que tiene. Esto es muy importante en la fidelización del cliente, por una parte para que identifique rápidamente los productos de su marca y por otra para transmitirle las ideas que quiere representar. Por esta razón se ha de seguir con esmero la guía de estilo de Movistar, incluso para un sistema de demostración como el que aquí se ha desarrollado.

Básicamente se han creado cinco tipos de elementos gráficos:

- **Icono principal de la aplicación.** Se ha diseñado un icono original para representar a la aplicación en el escritorio del PC servidor y de los terminales móviles. Se ha pretendido representar con un pequeño dibujo la idea de que hay una conferencia (con la aparición de los auriculares) y que participan más de dos personas (con tres bustos en distintos planos). Todo esto respetando la gama de colores y la simpleza de líneas de “Movistar”.
- **Iconos para botones.** Para ahondar en el aspecto visual y simple de la aplicación, se ha intentado incluir el menor texto posible. Por esto la práctica mayoría de los botones se han creado como iconos sin texto. Todos los iconos se han dibujado siguiendo la estética “Movistar”.
- **Fondo de la pantalla de inicio.** Se ha creado un fondo donde aparece el logo de “Movistar”, respetando el *look & feel* de la marca, que sirve de presentación de la aplicación ya que es la primera imagen en aparecer siempre al encenderla.

- **Fondos de sala.** Para ayudar al usuario a que se sienta inmerso en la sala virtual, se han creado unos fondos que representan las 3 situaciones básicas en que lo podrá estar usando: corporativo (con una mesa de reuniones), familiar (con sillones en un salón) y entre amigos (con ambiente informal). En todos los casos, el tamaño de los escenarios está pensado para aprovechar al máximo la colocación sobre ellos de los avatares. Para cada fondo existe una versión fotorrealista y otra como dibujo.
- **Avatares.** Son las imágenes que representan a los participantes. Suponen una parte imprescindible de la interfaz para que el usuario identifique inmediatamente a quién esté hablando y además tenga la libertad de colocarlo acústicamente en cualquier posición. Se ofrecen en modo foto o dibujo. Las fotos son de cada usuario posible, en primer plano para que se le reconozca bien, y les aplicamos esquinas redondeadas para suavizar la estética. Para los dibujos se permite elegir para cada usuario entre 12 posibles ilustraciones que mantienen el mismo estilo pictórico que los dibujos de los fondos. En los dos casos, se añade el nombre del participante debajo de cada avatar porque ayuda a su identificación.

Por último, indicar que para las ilustraciones de los escenarios de las salas y los avatares, hemos contado con la colaboración de una dibujante profesional por parte de Telefónica.



Figura 14. Icono principal de la aplicación.

5.3. Optimización de los algoritmos y tecnologías

Uno de los objetivos principales de este proyecto es la mejora de la calidad de la conversación en una multiconferencia para que tenga más realismo. Pero obviamente no queremos que esta mejora afecte seriamente al consumo de CPU del servidor y los terminales, ni al de la batería de los terminales, ni al de datos transmitidos por 3G. Por eso es muy importante la optimización de varias de las tecnologías que usamos.

En primer lugar tenemos el uso de **códec de compresión de voz Speex** que ha sido **adaptado y optimizado** para ser utilizado en teléfonos móviles inteligentes. Admite 3 calidades de audio: 8 kHz, 16 kHz y 32 kHz. Descartamos la más baja porque es el estándar en las redes móviles y da la sensación de sonido enlatado que queremos mejorar. Con 16 kHz la calidad de la voz mejora notablemente y es la calidad que hemos elegido. Descartamos 32 kHz porque la señal de voz nunca cubre tanto ancho de banda y no se puede apreciar un salto en calidad. El códec Speex tiene muchos más parámetros configurables que hacen variar la calidad a costa del coste computacional. Han sido necesarias numerosas pruebas para encontrar la solución de compromiso para todos estos parámetros.

En segundo lugar está la implementación del **algoritmo de espacialización del sonido basado en HRTF**. Existen muchas formas de llevar a la práctica este algoritmo teórico. Si tenemos en cuenta que en cada terminal se debe calcular para cada uno de los participantes presentes, nos damos cuenta que se debe aligerar al máximo su coste computacional. Así que esta es una de las partes que ha sido programada en C para ambas plataformas (no nos podemos permitir Java por ejemplo). Esta técnica binaural se fundamenta en aplicar una diferencia de retardo entre los oídos y unos filtros diferentes para cada oído según la dirección de llegada de la fuente. Es en el **diseño eficiente de estos filtros** donde hay más margen para ahorrar CPU y donde hemos hecho un esfuerzo especial.

En nuestro afán por **reducir el ancho de banda** necesario, también se han diseñado con cuidado los protocolos sobre IP para la transmisión de las señales:

- En el caso de la **señal de voz** comprimida, se ha decidido transmitir los paquetes haciendo uso de la **capa de transporte UDP**. El protocolo UDP es aconsejable para la transmisión de audio en tiempo real. No necesita establecer una conexión punto a punto, es más ágil y supone una menor sobrecarga de datos para la transmisión que TCP. El protocolo UDP no garantiza la recepción de los paquetes, pero así es preferible para audio en tiempo real, ya que si un paquete se pierde no nos sirve de nada que sea reenviado más tarde. Además, el códec Speex nos ayuda a rellenar "lo mejor posible" los espacios vacíos de los paquetes perdidos. Se ha creado una estructura especial para los paquetes que se van a enviar por UDP donde a los datos de audio se les va a añadir una cabecera muy pequeña con información sobre el usuario al que pertenece y número de secuencia. Como se podrá suponer, la identificación del usuario en la trama es un aspecto mucho más importante en una multiconferencia que en una simple llamada a dos.
- En el caso de los **mensajes de señalización** que se envían entre terminales y servidor para gestionar la multiconferencia, se ha decidido utilizar el **protocolo TCP**. Este tipo de mensajes son cortos y ocupan muy poco ancho de banda, pero

en este caso no nos podemos permitir perder algún paquete porque son vitales en la gestión de la comunicación. Además, establecer una conexión punto a punto entre servidor y terminales nos va a ayudar en este caso a monitorizar que cada punto sigue activo. Por todo esto el protocolo ideal es el TCP. Sobre este protocolo hemos diseñado nuestro propio tipo de trama que se adecúa a nuestras necesidades para gestionar una multiconferencia con dos tipos de mensajes: comandos o información.

Por último, se ha incluido un **detector de actividad de voz** que podemos activar opcionalmente. El detector de actividad analiza la señal de voz que captura el micrófono y nos indica el porcentaje de probabilidad de que haya voz en cada trama de audio. Gracias a esto podemos hacer que cuando no haya actividad de voz, la trama no se transmita. Para el terminal transmisor esto supone ahorrar el tiempo de compresión, pero beneficia todavía más a los terminales receptores porque no tendrán que descomprimir el paquete ni aplicarles la espacialización del sonido, usando así menos CPU. Además se ahorra el ancho de banda que ocupa la distribución de esa señal de voz entre todos los terminales. Utilizar el detector de actividad tiene más beneficios cuantos más participantes hay en la multiconferencia, ya que seguramente serán muchos de ellos los que se encuentren en silencio en un momento dado.

5.4. Implementación de la aplicación para los terminales

Prototipo de la aplicación terminal en un PC

A la hora de implementar las aplicaciones para los terminales, se ha tenido en cuenta que la programación para las plataformas móviles, a pesar de contar con herramientas avanzadas, es más lenta y requiere código más largo y complejo que con otras herramientas y lenguajes existentes en la programación para PC (bajo Windows por ejemplo).

Por esta razón se decidió programar un prototipo de la aplicación cliente de multiconferencia para PC en lenguaje Delphi. Se eligió este lenguaje por ser uno de los que permiten mayor velocidad de desarrollo, de forma que con el mínimo código posible se puede implementar y probar cualquier idea en tiempo record, incluido la creación de complejas interfaces gráficas. Esta decisión supuso un gran acierto y ahorró muchas horas de trabajo si consideramos que existían muchas partes en el programa para las que había que probar distintas soluciones, y con el prototipo de PC se podían evaluar todas en una fracción de tiempo. Así mismo, el prototipo versión PC nos permitía mostrar pruebas de concepto a distancia a nuestro cliente Telefónica para que realizara un seguimiento del proyecto y diera el visto bueno a las características más complicadas.

Con esta filosofía de trabajo, una vez quedaba validado cada detalle final de la interfaz gráfica o del procesado de audio en el prototipo para PC, se trasladaba el código a las plataformas móviles, con sus respectivas características.

Lenguaje de programación para los terminales

En cuanto al tipo de código para los terminales, se decidió implementar en C todo el que tenía que ver el procesado de la señal, porque se tratan de operaciones en cierto modo independientes al tipo de plataforma móvil y porque es imprescindible que se ejecuten con máxima eficiencia, algo que no pueden garantizar los lenguajes que colocan estos sistemas operativos como capa superior a C (Objective C para iOS y Java para Android). Aun así, la forma de tratar con código nativo en C es distinta en cada una de estas plataformas, por lo que este código solo es parcialmente reutilizable.

La programación de las interfaces gráficas, en cambio, son muy dependientes de la plataforma, y la única forma eficiente de implementarlas es utilizando las librerías propias que ofrece cada SDK, sobre Objective-C en iOS [31] y Java en Android [32]. La filosofía de programación es muy distinta en cada plataforma, por lo que es un reto intentar obtener cierto grado de homogeneización entre ambas.

Resoluciones y relaciones de aspecto

Uno de puntos laboriosos de la programación de estas interfaces es el hecho de que los terminales pueden tener distintas resoluciones y relaciones de aspecto. Esto implica por una parte crear los elementos artísticos en distintas resoluciones y por otra colocar los objetos en las posiciones adecuadas para que este hecho sea transparente al usuario:

- **En el caso de iOS**, las posibles resoluciones están en cierto modo controladas porque solo hay dos tipos de dispositivos: los iPad que tienen 1024x768 píxeles de resolución y los iPhone que pueden tener 480x320 o 960x640 píxeles de resolución según el modelo. Si nos damos cuenta, en los iPhone las resoluciones posibles son una el doble exacta de la otra, lo que facilita las cosas. Se han creado 3 resoluciones para todos los elementos gráficos y es el propio sistema operativo quien elige el adecuado según el modelo sobre el que se esté ejecutando. Dado que la relación de aspecto no es la misma en iPhone que en iPad, algunos detalles estéticos son necesarios programarlos de forma distinta para iPhone que para iPad.
- **En el caso de Android**, dado que es utilizado en muchísimas marcas y modelos de terminales móviles, nos podemos encontrar pantallas con muy distintas

resoluciones y relaciones de aspecto. En principio esto es algo que la plataforma ha solucionado obligando a que los programadores utilicen coordenadas y proporciones como datos relativos a la pantalla y no absolutos. Y aunque esta filosofía funciona bastante bien, es bastante difícil que las interfaces queden siempre tan bien cuidadas en cualquier terminal. En este proyecto piloto, que no va a ser comercializado de momento para todos los terminales, nos hemos asegurado que ciertos modelos que con seguridad se van a utilizar en las demostraciones, tengan una interfaz cuidada con precisión, homogénea con la de iOS. Para el resto de modelos que no hemos podido probar (porque no es asumible hacerlo para un proyecto de este tipo), la interfaz puede ser algo menos precisa o parecida en sus proporciones, pero mantiene igualmente una buena calidad.



Figura 15. Pantallas de inicio, principal y diferentes menús en iOS y Android respectivamente.

5.5. Implementación del servidor de multiconferencia

El servidor se ha programado pensando en las **necesidades actuales del sistema piloto**, es decir, para que sea cómodo para cualquier encargado de organizar una demostración, para que proporcione información útil de cómo utilizan los usuarios el sistema y para que sea robusto considerando que esta versión no será utilizada por miles de usuarios a la vez pero que una vez encendido no tiene por qué apagarse en mucho tiempo.

Si bien por los requerimientos que en un principio se nos hicieron, el servidor gestiona una única sala de multiconferencia a la vez, se ha dejado preparado para que sea ampliable a varias salas cuando así se requiera más adelante en otras demostraciones. Aunque una futura versión comercial escalable a miles de usuarios y salas simultáneas requeriría la reprogramación del servidor (al ser las condiciones y objetivos distintos), todas las técnicas que se han aplicado a este servidor han sido pensadas por su utilidad en una escalabilidad futura y minimizan el uso de CPU.

Estas son las **principales funciones** que realiza el servidor:

- Guarda el estado en que se encuentra la sala de multiconferencia, con la información de todos los usuarios que están dentro, pudiendo limitar si se desea el número máximo de participantes.
- Cuando un terminal le solicita entrar en la sala de multiconferencia, el servidor comprueba en su base de datos que el usuario esté dado de alta, que la contraseña sea correcta y que no haya otro terminal usando el mismo usuario. Si es todo correcto, envía al resto de participantes de la sala la información del nuevo participante que ha entrado para que aparezca en todos los terminales.
- Cuando un terminal le solicita abandonar la conversación, el servidor lo informa al resto de participantes y lo elimina de la sala.
- Recibe el flujo de voz de cada participante de la sala y lo retransmite replicándolo a todo el resto de participantes en la sala.
- Proporciona una salida textual con información de la gestión que está realizando el servidor en cada instante.

A nivel de conectividad, se ha programado el servidor para que solo utilice dos puertos, uno para los mensajes TCP y otro para los flujos de audio UDP. Era importante reducir a solo dos puertos los necesarios ya que el servidor se va a conectar dentro de todo tipo de redes institucionales o empresariales, y a menudo la solicitud para que el administrador de algunas redes abra ciertos puertos de entrada es complicada. Se han resuelto a nivel de programación también otros problemas derivados del uso de routers intermedios que podrían provocar que una conexión no encontrara su ruta entre los dos extremos cuando el servidor o los terminales no se están conectados a la red mediante una IP pública sino local.

Interfaz del servidor

De cara al encargado de llevar una demostración, cuando abra la aplicación servidor verá directamente la pantalla principal donde tocando el botón de encendido el servidor ya estará funcionando.



Figura 16. Pantalla principal del servidor de multiconferencia.

Existe una segunda pantalla donde se puede configurar algún ajuste (normalmente no es necesario) y donde se muestra la información del estado del servidor y lo que está sucediendo en la sala.

Capítulo 6

Resultados y trabajo futuro

6.1. Resultados obtenidos

En este proyecto se ha desarrollado un **innovador sistema de multiconferencia móvil con audio en 3D**. Como se ha explicado en esta tesina, es una unión novedosa, de la cual no hay constancia que exista ninguna parecida en el mercado, de cuatro tecnologías ya existentes:

- La voz sobre IP en teléfonos móviles.
- Los servicios de multiconferencia.
- Las técnicas de audio 3D de gran realismo mediante HRTF .
- Las interfaces gráficas sobre pantallas táctiles en dispositivos móviles.

Esta combinación se ha realizado con el propósito de **mejorar sustancialmente la inteligibilidad y la identificación de los participantes** en una multiconferencia móvil.

Además, para implementar con éxito este producto, se ha hecho un importante trabajo de **optimización y adaptación de estas técnicas** ya conocidas en otros ámbitos para que funcionen adecuadamente bajo estas condiciones.

Para dar una idea del volumen de trabajo que ha supuesto este proyecto podemos aportar algunos datos:

- El presente autor ha dedicado **6 meses, más de 600 horas**, en el desarrollo de este proyecto, bajo la dirección muy implicada de su jefe/tutor en el iTEAM y esta tesina.
- Se han programado **3 aplicaciones** (servidor y terminales iOS y Android) con un total aproximado de **10,000 líneas de código** (2,000 para el servidor, 5,000 para iOS y 3,000 para Android).

- Se ha generado **numerosa documentación** interna para Telefónica incluyendo propuestas técnicas, manuales de uso, informes técnicos y presentaciones.
- Se han mantenido unas **15 reuniones** con el contratante Telefónica: 4 presenciales en su sede central de Madrid y el resto por vía telefónica (alguna incluso a través de este propio sistema de multiconferencia).
- Se ha acudido y dado soporte técnico en la demostración de este proyecto durante tres días en el *Mobile World Congress 2011* de Barcelona .

El sistema desarrollado ha sido intensamente probado, tanto internamente en los laboratorios del iTEAM, como por el personal de Telefónica, para verificar que se han alcanzado los objetivos propuestos en cuanto a robustez del sistema, facilidad de uso y experiencia auditiva. Gracias a los protocolos simples pero eficientes que se han implementado, se ha logrado un sistema muy fiable que se ha comprobado que no se bloquea salvo en situaciones extremas.

Los resultados obtenidos han sido totalmente satisfactorios según Telefónica y según terceros profesionales de la prensa especializados que han probado este sistema de multiconferencia.

Resumiendo, se ha logrado que el usuario de este nuevo servicio tenga la sensación de estar reunido en una sala con otros participantes de una forma muchísimo mas natural y real de lo que hasta ahora los servicios existentes nos han ofrecido.

6.2. Repercusión del proyecto

Este sistema piloto de multiconferencia, ya desde antes de estar finalizado del todo, ha sido sometido a diversas **demostraciones internas en Telefónica**. Varios departamentos de la empresa han podido probar este nuevo servicio. Desde el iTEAM hemos tenido la oportunidad de hablar con distinto personal de esta operadora de telecomunicaciones que nos han manifestado la **buena acogida** interna que ha tenido y las posibilidades que le veían.

Pero como es obvio, desde Telefónica necesitan comprobar la reacción que provoca este servicio piloto **fuera de la empresa ante posibles clientes y ante otros profesionales del sector**. Para ello han intentado dar a conocer este sistema en distintos lugares, donde podríamos destacar su exposición en el *Mobile World Congress 2011* de Barcelona y en el *SICARM 2011* de Murcia.

6.2.1. Mobile World Congress 2011 Barcelona

El *Mobile World Congress* (MWC) es el **mayor congreso anual** sobre las comunicaciones móviles a nivel mundial. En la edición del 2011, celebrada en Barcelona, se dieron cita unas 1300 compañías haciendo una exhibición mundial de la industria móvil. Bajo este marco de tendencias se reunieron los mayores y más innovadores líderes y personalidades de los operadores y proveedores de la industria móvil [33].

Como en anteriores ediciones, **Telefónica dispuso en la feria de un gran stand** (367 m2 en dos alturas) para dar a conocer su visión de futuro de la industria móvil en cuanto a servicios y tecnologías [34]. En dicho stand expusieron varios productos y servicios novedosos que están estudiando, entre ellos este sistema de multiconferencia que hemos desarrollado, al que se le dio el nombre corto de **“Multiconferencia Inmersiva”** [35].

Este proyecto fue mostrado tanto al público asistente al congreso como un grupo de periodistas especializados, ofreciéndoles una rueda de prensa expresamente para explicarles este proyecto. En todos los casos **fue una gran sensación**. Así mismo, fue uno de los dos proyectos que Telefónica decidió que probaran el grupo de autoridades españolas que visitó el stand para inaugurarlos.



Figura 17. Autoridades probando la “Multiconferencia Inmersiva” en el MWC 2011. De izquierda a derecha: Jordi Hereu (alcalde de Barcelona), Miguel Sebastián (Ministro de Industria, Turismo y Comercio), Artur Mas (presidente de la Generalitat de Catalunya) y Julio Linares (consejero delegado de Telefónica S.A.)

6.2.2. SICARM 2011 Murcia

El SICARM es el foro anual de la Sociedad de la Información de la Región de Murcia, formado por un conjunto de eventos tecnológicos donde además de analizar y experimentar las “nuevas tecnologías” desde diversas perspectivas, se presentan las últimas novedades en los sectores de la Informática y las Telecomunicaciones [36].

En la edición del 2011 participó Telefónica presentando los nuevos productos y servicios que están evaluando. Entre otros presentó este proyecto de “Multiconferencia Inmersiva” para darlo a conocer también allí. Tuvo una buena acogida y también fue recogido por la prensa.

6.2.3. Repercusión en la prensa

Esta plataforma piloto de multiconferencia con audio espacial para terminales móviles, bajo el nombre más corto y apropiado para los medios de “Multiconferencia Inmersiva”, ya ha logrado aparecer en diversas ocasiones en **noticias tecnológicas de la prensa.**

Los dos ferias tecnológicas nombradas anteriormente (MWC y SIRCAM) han potenciado estas apariciones gracias también a las ruedas de prensa ofrecidas allí por Telefónica acerca de sus novedades.

En algunos casos se han escrito artículos o noticias enteras comentando las bondades de este servicio de “Multiconferencia Inmersiva” y en otros solo aparece nombrado o explicado en una parte de la publicación.

Algunos medios relevantes que se han hecho eco de este proyecto son **Europa Press** [37], **El Mundo** [38], **La Razón** [39], **ABC** [40], **Cinco Días** [41], **El Economista** [42], **PC Actual** [43] y otros [44] [45] [46].

Algunos de los titulares que han encabezado estas noticias:

- “Telefónica quiere hacer de la multiconferencia una experiencia más real” – *Europa Press* [37]
- “Telefónica desarrolla una multiconferencia con sonido en 3D” – *ABC* [40]
- “Telefónica lleva la multiconferencia inmersiva al móvil” – *Network World* [47]
- “Telefónica presenta Immersive Conference, VoIP con sensación 3D” – *XatakaON* [48]
- “Telefónica quiere mejorar la experiencia de las multiconferencias” – *Nuevas Tecnologías* [49]

6.3. Trabajo futuro

Una vez finalizado este proyecto, y después de que Telefónica haya hecho numerosas demostraciones y pruebas en los sitios con más interés estratégico, ha surgido la posibilidad de continuar avanzando en esta línea con nuevas fases. A corto plazo se van a incorporar nuevas funcionalidades en el demostrador para continuar su estudio. A medio y largo plazo Telefónica está evaluando la posibilidad de lanzar una versión comercial de este servicio.

6.3.1. Mejoras del demostrador

Cuando se diseñó esta plataforma piloto, se contemplaba su uso en entornos muy controlados, de forma que al realizar una demostración se sabe de antemano quienes van a participar y se prepara todo para tal efecto. Esto tenía que ser así porque el objetivo era principalmente validar la experiencia de usuario ante las nuevas técnicas de audio 3D junto con su interfaz táctil. Es decir, el objeto de estudio era la experiencia de cada individuo en concreto.

Una vez logrado este objetivo, ahora se pretende mejorar el demostrador para que se use de un modo más libre, con menos control del entorno. Esto quiere decir que el objeto de estudio pasa a ser el comportamiento global y a largo plazo de los potenciales usuarios. Las demostraciones por tanto van a pasar de ser en momentos puntuales a ser continuas en el tiempo, dejando que ciertos grupos de usuarios vayan haciendo uso de este sistema en casos más reales, es decir, en cualquier momento en que lo vayan necesitando.

Las mejoras que se han propuesto hacer a tal efecto tienen que ver con **dar más autonomía al servidor** para que no tenga que haber un encargado organizando las demostraciones sino que el servicio tenga más vida propia. Destacamos dos nuevas funcionalidades necesarias entre otras menores:

- Dado que no se va a ordenar y organizar el uso, es imprescindible que el **servidor gestione simultáneamente distintas salas de multiconferencia**, con los cambios que ello implica en el servidor (que ya está parcialmente preparado) y en la interfaz de los terminales (añadiendo algunas opciones).
- Lo que interesa ahora es poder analizar el comportamiento de los usuarios a largo de los días. Por eso se hace necesaria la incorporación de un **módulo de estadísticas que recopile información útil** sobre como se está utilizando el servicio (valores medios como cuantos clientes se conectan al servicio, cuanto tiempo hablan, cuantas salas utilizan, etc). De este modo, posteriormente, se podrá consultar el registro para analizar como está siendo usado y obtener conclusiones que permitan continuar mejorando el servicio.

6.3.2. Posibilidades de comercialización

Las grandes operadoras de comunicaciones, como Telefónica, no cesan de buscar innovaciones tecnológicas que puedan convertirse en nuevas líneas de negocios. Para ello continuamente están probando una gran cantidad de tecnologías y prototipos que consideran interesantes. Como es de suponer, un porcentaje muy bajo de todo lo que evalúan acaba convirtiéndose en un producto comercial.

En estos momentos, la plataforma piloto desarrollada sigue siendo probada y analizada en Telefónica, y esperan continuar su evaluación con las nuevas mejoras que han solicitado. El proceso para que pueda llegar al mercado es lento y largo. Deben evaluar todo tipo de pros y contras, y ya no solo a nivel técnico, si no también a nivel comercial.

¿Existe una demanda en el mercado para un servicio así? ¿Qué inversión supondría? ¿Cuál es el modelo de negocio que habría que aplicar? ¿Cómo se facturaría? Todas estas son preguntas que tienen resolver los expertos de los departamentos comerciales de una operadora de comunicaciones. Como instituto de investigación tecnológica, estos aspectos no caen en nuestras manos. La evolución futura a nivel comercial es por tanto incierta para nosotros, no sabemos hasta donde puede llegar, solo que de momento continúa siendo evaluada.

Bibliografía

- [1] Página web de Telefónica: <http://www.telefonica.com/>
- [2] Página web del iTEAM: <http://www.iteam.upv.es/>
- [3] Página web del GTAC: <http://www.gtac.upv.es/>
- [4] A.Bronkhorst. "The Cocktail Party Phenomenon: A Review on Speech Intelligibility in Multiple-Talker Conditions". *Acta Acustica united with Acustica*, 2000, 86: 117–128.
- [5] Página web de GSM: <http://www.gsm.org/>
- [6] C. Christopher, "Essential of UMTS". *Cambridge University Press*, 2008.
- [7] S. Ahson, M. Ilyas, "The VoIP Handbook". *CRC Press*, 2008.
- [8] Página web de iOS: <http://www.apple.com/es/ios/>
- [9] Página web de Android: <http://www.android.com/>
- [10] Página web de WP7: <http://www.microsoft.com/windowsphone/>
- [11] Página web del Blackberry: <http://www.blackberry.com/>
- [12] Página web de Symbian: <http://symbian.nokia.com/>
- [13] Página web de Bada: <http://www.bada.com/>
- [14] Informe de Gartner sobre ventas de terminales móviles en el primer trimestre del 2011: <http://www.gartner.com/it/page.jsp?id=1689814>
- [15] Página web de Movistar Empresas <http://www.movistar.es/empresas>
- [16] Página web de PowWowNow: <http://www.powwownow.es/>

- [17] Página web de BTMeetMe: <http://www.meetme.bt.com/>
- [18] Página web de Arkadin: <http://www.arkadin.es/>
- [19] Página web de Mirial ClearSea: <http://www.mirial.com/>
- [20] Página web de Skype: <http://www.skype.com>
- [21] J. Engdegard et al., “Spatial Audio Object Coding (SAOC) – The Upcoming MPEG Standard on Parametric Object Based Audio Coding”, *AES Convention Paper*, 2008.
- [22] J. M. Valin, “Speex: A Free Codec For Free Speech”, *linux.conf.au*, 2006.
- [23] J. M. Valin, “The Speex Codec Manual Version 1.2 Beta 3”, 2007.
- [24] Página web de Speex: <http://www.speex.org/>
- [25] D. Begault. “3D Sound for Virtual Reality and Multimedia”, *A. Press*, 1994.
- [26] J. Blauert, “Spatial Hearing: the psychophysics of human sound localization”, *MIT Press*, 1997.
- [27] Página web de iOS Dev Center:
<http://developer.apple.com/devcenter/ios/index.action>
- [28] Página web de iOS Developer Program:
<http://developer.apple.com/programs/ios/>
- [29] Página web de Android Developers:<http://developer.android.com/index.html>
- [30] Página web de Android Market: <https://market.android.com/>
- [31] M. Ali, “Advanced iOS 4 Programming”. *Wiley*, 2010.
- [32] J. Steele, N. To, “The Android developer’s cookbook”. *Addison-Wesley*, 2010.
- [33] Página web del Mobile World Congress:
<http://www.mobileworldcongress.com/>
- [34] Página web de Telefónica en el MWC 2011:
<http://www.telefonica.com/es/mwc/home/index.shtml>
- [35] Briefing de prensa “Multiconferencia Inmersiva” de Telefónica MWC 2011:
http://www.telefonica.com/es/descargas/mwc/Multiconferencia_Inmersiva_briefing_prensa.pdf

- [36] Página web de la feria SICARM: <http://www.sicarm.es/>
- [37] Noticia en Europa Press: <http://www.europapress.es/economia/noticia-economia-telecos-telefonica-quiere-hacer-multiconferencia-experiencia-mas-real-20110215161650.html>
- [38] Noticia en El Mundo: <http://www.elmundo.es/elmundo/2011/02/15/navegante/1297798628.html>
- [39] Noticia en La Razón: <http://www.larazon.es/noticia/5655-con-el-dni-pegado-al-movil>
- [40] Noticia en ABC: <http://www.abc.es/20110215/tecnologia/abci-telefonica-desarrolla-multiconferencia-sonido-201102151723.html>
- [41] Noticia en Cinco Días: http://www.cincodias.com/articulo/empresas/telecos-consideran-caro-pagar-2000-millones-espectro/20110215cdscdiemp_1/
- [42] Noticia en El Economista: <http://www.economista.es/economia/noticias/2828662/02/11/Telefonica-presenta-la-multiconferencia-con-sonido-3d.html>
- [43] Noticia en PC Actual: http://www.pcactual.com/articulo/actualidad/noticias/mobile_world_congress_2011/7937/mwc_movistar_muestra_las_aplicaciones_del_futuro.html
- [44] Noticia en Terra: <http://www.terra.es/tecnologia/lared/articulo/telefonica-presenta-multiconferencia-inmersiva-27180.htm>
- [45] Noticia en Invertia: <http://www.invertia.com/noticias/articulo-final.asp?idNoticia=2475333>
- [46] Noticia en la Opinión de Murcia: <http://www.laopiniondemurcia.es/comunidad/2011/04/14/ordenadores-toman-tension/316268.html>
- [47] Noticia en Network World: <http://www.networkworld.es/Telefonica-lleva-la-multiconferencia-inmersiva-al-/seccion-actualidad/noticia-106297>
- [48] Noticia en XatakaOn: <http://www.xatakaon.com/tecnologia-voip/telefonica-presenta-immersive-conference-voip-con-sensacion-de-3d>
- [49] Noticia en Nuevas Tecnologías: <http://www.nuevastecnologias.com/telefonica-quiere-mejorar-la-experiencia-de-las-multiconferencias-15-02-2011/>