

Análisis del Uso de Interactividad basada en Voz para el manejo del Servicio de Video bajo Demanda

LAURA GÓMEZ FIGUEROA
JESÚS ARMANDO SANDOVAL LEÓN

Director
Ing. OSCAR JOSUÉ CALDERÓN CÓRTEZ

Trabajo de grado como requisito para obtener el título de Ingeniero en Electrónica
y Telecomunicaciones

UNIVERSIDAD DEL CAUCA
FACULTAD DE INGENIERÍA ELECTRÓNICA Y
TELECOMUNICACIONES
DEPARTAMENTO DE TELECOMUNICACIONES
GRUPO I+D EN NUEVAS TECNOLOGÍAS EN
TELECOMUNICACIONES

Análisis del Uso de Interactividad basada en Voz
para el manejo del Servicio de Video bajo
Demanda



LAURA GÓMEZ FIGUEROA
JESÚS ARMANDO SANDOVAL LEÓN

UNIVERSIDAD DEL CAUCA
FACULTAD DE INGENIERÍA ELECTRÓNICA Y
TELECOMUNICACIONES
DEPARTAMENTO DE TELECOMUNICACIONES
GRUPO I+D EN NUEVAS TECNOLOGÍAS EN TELECOMUNICACIONES

Análisis del Uso de Interactividad basada en Voz
para el manejo del Servicio de Video bajo
Demanda



LAURA GÓMEZ FIGUEROA
JESÚS ARMANDO SANDOVAL LEÓN

DIR. Ing. OSCAR JOSUÉ CALDERÓN CORTÉS

Trabajo de grado como requisito para obtener el título de Ingeniero en Electrónica y
Telecomunicaciones

UNIVERSIDAD DEL CAUCA
FACULTAD DE INGENIERÍA ELECTRÓNICA Y
TELECOMUNICACIONES
DEPARTAMENTO DE TELECOMUNICACIONES
GRUPO I+D EN NUEVAS TECNOLOGÍAS EN TELECOMUNICACIONES

Agradezco y dedico este trabajo
A la mujer que por tantos años me brindó su apoyo y guió mi
vida, mi ABUELITA

A mi familia por su apoyo incondicional
A mis amigos por tantos momentos vividos

LAURA GÓMEZ FIGUEROA

A Dios, por permitirme llegar a este momento tan especial en mi vida. Por los triunfos y los momentos difíciles que me han enseñado a amarte cada día más.

A mi familia que siempre estuvo ahí brindándome su apoyo en todo momento.

A mi madre Orleida y mi padre Enrique por su constante apoyo, por sus consejos, por los ejemplos de perseverancia que siempre los caracterizan y que me han infundado día a día. Va por ustedes, por lo que valen, porque admiro su fortaleza y por lo que han hecho de mí.

A mis hermano y a mi hermana porque siempre he contado con ellos para todo, gracias a la confianza que siempre nos hemos tenido.

A mis amigos y compañeros que nos apoyamos mutuamente en nuestra formación profesional y que hasta ahora, seguimos siendo amigos.

A mi compañera de trabajo de grado, gracias al equipo que formamos logramos llegar hasta el final del camino.

A mis maestros Gracias por su tiempo, por su apoyo, en especial: al Ingeniero Oscar Calderon por haber guiado el desarrollo de este trabajo y llegar a la culminación del mismo.

A la Universidad del Cauca, y en especial a la facultad de Ingeniería en Electrónica y Telecomunicaciones, por permitirme ser parte de una generación de triunfadores y gente productiva para el país.

JESÚS ARMANDO SANDOVAL LEÓN

AGRADECIMIENTOS

Al ingeniero Oscar Josué Calderón Cortés por su apoyo, dedicación y comprensión durante el desarrollo del proyecto y los años de estudiantes universitarios. Al ingeniero Guefry Agredo por su apoyo y colaboración. Al ingeniero Victor Manuel Mondragón por la colaboración dentro del proyecto.

TABLA DE CONTENIDO

<i>INTRODUCCIÓN</i>	1
<i>CAPÍTULO 1. LA TELEVISIÓN Y LOS MECANISMOS DE INTERACTIVIDAD</i>	3
1.1 INTRODUCCIÓN	3
1.2 LA TELEVISIÓN	3
1.2.1 Definición	4
1.2.2 Interactividad.....	4
1.2.3 Televisión Interactiva	5
1.3 MECANISMOS DE INTERACTIVIDAD EN UN AMBIENTE DE DISTRIBUCIÓN DE TELEVISIÓN	8
1.3.1 Principales mecanismos de Interacción en un ambiente de televisión	8
1.3.1.1 El Control Remoto.....	9
1.3.1.2 La Voz.....	10
1.3.1.3 Interfaz Táctil	10
1.3.1.4 Interfaz Gestual	11
1.3.1.5 Interfaces Multimodales	12
1.3.1.6 Otras Interfaces	12
1.3.2 Clasificación de los Mecanismos de Interactividad.....	15
1.3.2.1 Por las Acciones a Desarrollar	15
1.3.2.2 Según la Población Objetivo.....	16
1.3.2.3 Por la Tecnologías Involucradas	17
<i>CAPÍTULO 2. CONCEPTOS ESENCIALES DE UN ESCENARIO DE SERVICIO DE VIDEO BAJO DEMANDA</i>	19
2.1 INTRODUCCIÓN	19
2.2 FUNDAMENTOS DE VIDEO BAJO DEMANDA	19
2.2.1 Definición	19

2.3	ESCENARIO DE SERVICIO DE VOD.....	23
2.3.1	Servidores.....	23
2.3.2	Red de Comunicaciones	26
2.3.3	Protocolos de Red	28
2.3.3.1	Capa de Codificación.....	28
2.3.3.2	Capa de Empaquetamiento	28
2.3.3.3	Capa de Construcción del Flujo de transporte	29
2.3.3.4	Protocolo de Transporte en Tiempo Real.....	29
2.3.3.5	Protocolo de Flujo de Tiempo Real	29
2.3.3.6	Protocolo de Control de Tiempo Real.....	29
2.3.3.7	Protocolo de Control de Transmisión.....	30
2.3.3.8	Protocolo de Datagramas de Usuario	30
2.3.3.9	Protocolo de Descripción de Sesión	30
2.3.3.10	Protocolo de Internet	31
2.3.3.11	Capa de Enlace.....	31
2.3.3.12	Capa Física	31
2.3.4	Terminal de Usuario	31
2.3.4.1	Set Top Box	32
2.3.4.2	Centro Multimedia.....	32
2.3.5	Proceso de Petición y Entrega	33

***CAPÍTULO 3. PROPUESTA DE UNA SOLUCIÓN PILOTO DE INTERACTIVIDAD
BASADA EN VOZ SOBRE IP PARA EL MANEJO DEL SERVICIO DE VIDEO BAJO
DEMANDA.....*** **35**

3.1	INTRODUCCIÓN.....	35
3.2	PROCESO DE DESARROLLO.....	35
3.3	ESTABLECIMIENTO DE LOS REQUISITOS DE LA SOLUCIÓN PILOTO.....	37
3.4	GENERACIÓN Y DESCRIPCIÓN GENERAL DE LA SOLUCIÓN	38
3.5	IMPLEMENTACIÓN, CONFIGURACIÓN Y PUESTA A PUNTO	40
3.5.1	Actividad 1: Configuración de los Servidores de Video y Web	41
3.5.1.1	Servidor de Video Streaming	41
3.5.1.2	Servidor Web	43
3.5.2	Actividad 2: Instalación y Configuración del Servidor de Comunicaciones IP y el Motor de Reconocimiento.....	45

3.5.2.1	Servidor de Comunicaciones IP	45
3.5.2.2	Motor de Reconocimiento	50
3.5.3	Actividad 3: Despliegue del Cliente e Integración del Sistema	51
3.5.3.1	Cliente VoIP	51
3.5.3.2	Cliente Multimedia.....	52
3.6	CONTEXTO DE USO DE LA SOLUCIÓN	57
3.6.1	Contexto de Uso de la Solución.....	57
3.6.2	Condiciones para la Aplicación de la Solución en un Entorno Real	58
 <i>CAPÍTULO 4. EVALUACIÓN DE LA SOLUCIÓN PILOTO Y ANÁLISIS DE RESULTADOS.....</i>		 61
4.1	ASPECTOS A EVALUAR	61
4.1.1	Enfoque Subjetivo	61
4.1.2	Enfoque Objetivo	63
4.2	ESCENARIO DE EVALUACIÓN.....	63
4.2.1	Entorno de Evaluación	63
4.2.2	Descripción de la Población	65
4.3	EVALUACIÓN SUBJETIVA	65
4.3.1	Proceso de Evaluación	65
4.3.2	Aplicación de la Evaluación	66
4.3.2.1	Introducción.....	66
4.3.2.2	Interacción con el Sistema con control remoto.....	66
4.3.2.3	Interacción con la Solución Propuesta.....	67
4.3.3	Análisis de Resultados	67
4.3.3.1	Atractivo del Sistema.....	67
4.3.3.2	Grado de Control.....	69
4.3.3.3	Eficiencia.....	71
4.3.3.4	Utilidad	73
4.3.3.5	Facilidad de Aprendizaje	74
4.4	EVALUACIÓN OBJETIVA	76
4.4.1	Comportamiento General de la Solución.....	76
4.4.2	Comportamiento del Flujo de Video.....	80
4.4.3	Comportamiento de la Comunicación VoIP	80
4.4.4	Comportamiento Socket TCP	81

4.4.5	Retardo Total.....	82
<i>Capítulo 5. CONCLUSIONES, RECOMENDACIONES Y TRABAJO FUTURO.....</i>		84
5.1	CONCLUSIONES	84
5.2	RECOMENDACIONES.....	85
5.3	TRABAJO FUTURO	85
REFERENCIAS.....		87

INDICE DE TABLAS

Tabla 1. Niveles de Interactividad de Acuerdo a las acciones a realizar.	6
Tabla 2. Niveles de Interactividad de acuerdo al servicio.....	6
Tabla 3. Niveles de Interactividad de acuerdo a las acciones y los servicios.	7
Tabla 4. Tipos de interactividad dependiendo del lugar de ejecución de los comandos.....	8
Tabla 5. Principales Mecanismos de Interacción.....	14
Tabla 6. Clasificación según las Acciones a Desarrollar.	15
Tabla 7. Clasificación de acuerdo a los usuarios con problemas físicos.	17
Tabla 8. Clasificación de acuerdo a las tecnologías involucradas.	17
Tabla 9. Tipos de VoD.....	21
Tabla 10. Arquitecturas de Organización de Servidores.	22
Tabla 11. Códec de Audio en un entorno IP.....	24
Tabla 12. Códec de Video.	25
Tabla 13. Medios de Transmisión usados en la red de distribución y acceso.	27
Tabla 14. Servidores de Video Streaming bajo licencia privativa.....	41
Tabla 15. Servidores de Video Streaming bajo licencia libre.....	42
Tabla 16. Servidores web disponibles actualmente.	44
Tabla 17. Servidores de comunicaciones IP disponibles actualmente.....	46
Tabla 18. Protocolos de señalización asociados a VoIP.	48
Tabla 19. Códec Asociados a VoIP soportados por Asterisk.....	48
Tabla 20. Motores de Reconocimiento.	50
Tabla 21. Comportamiento del Flujo de Video.	80
Tabla 22. Comportamiento de la Comunicación de VoIP.....	81

INDICE DE FIGURAS

Figura 1. Interfaz táctil asociada al control remoto.	11
Figura 2. Interacción mediante un cubo.	13
Figura 3. Escenario de Servicio de VoD.....	23
Figura 4. Pila de Protocolos.	28
Figura 5. Proceso de Petición y Entrega.....	33
Figura 6. Fases de desarrollo del proyecto.....	36
Figura 7. Actividades de la Fase de Implementación, Configuración y Puesta a Punto.	37
Figura 8. Esquema modular de una solución de interactividad.	40
Figura 9. Interfaz Gráfica de Usuario.....	52
Figura 10. Solución de Interactividad.....	54
Figura 11. Proceso de Interacción.....	55
Figura 12. Infraestructura de red de Imagenio.	58
Figura 13. Comportamiento del motor de reconocimiento de LumenVox.....	60
Figura 14. Escenario de Evaluación de la solución.....	64
Figura 15. Escenario de Evaluación con el Control Remoto.....	65
Figura 16. Proceso de evaluación.....	66
Figura 17. Interfaz de Servicio de VoD a través del Control Remoto.	67
Figura 18. Atractivo del Sistema Controlado por Control Remoto.	68
Figura 19. Atractivo del Sistema Controlado por Voz.	68
Figura 20. Grado de Control del Sistema con Control Remoto.	70
Figura 21. Grado de Control del Sistema con control por Voz.	70
Figura 22. Eficiencia del sistema con Control Remoto.	72
Figura 23. Eficiencia del Sistema con control por Voz.....	72
Figura 24. Utilidad Percibida del Sistema manejado con Control Remoto.	73
Figura 25. Utilidad Percibida del Sistema manejado por Voz.....	74
Figura 26. Facilidad de Aprendizaje del Sistema manejado por Control Remoto.	75

Figura 27. Facilidad de Aprendizaje del Sistema Manejado por Voz.	75
Figura 28. Tráfico total de la solución piloto.	78
Figura 29. Acercamiento al comportamiento del tráfico de voz, HTTP y TCP.	79
Figura 30. Ancho de banda consumido por el códec G.711.	81
Figura 31. Diagrama de Flujo de los paquetes TCP.	82

ACRÓNIMOS

AAC	Codificación Avanzada de Audio - Advanced Audio Coding
ADPCM	Modulación por Pulsos Codificados Diferencial Adaptativo – Adaptive Differential Pulse-Code Modulation
AGI	Interfaz a la Pasarela Asterisk - Asterisk Gateway Interface
ASR	Reconocimiento Automático del Habla – Automatic Speech Recognition
AVC	Códec Avanzado de Video - Advanced Video Codec
BBC	Corporación de Radiodifusión Británica - British Broadcasting Corporation
CPS	Cuestionario de Percepción del Servicio
DRM	Gestión de Derechos Digitales - Digital Right Management
DSL	Línea de Abonado Digital - Digital Subscriber Line
DSS	Darwin Streaming Server
EPG	Guía de Programación Electrónica – Electronic Program Guide
FIET	Facultad de Ingeniería Electrónica y Telecomunicaciones
GUI	Interfaz Gráfica de Usuario – Graphical User Interface
HD	Alta Definición – High Definition
IAX	Inter-Asterisk eXchange
IEEE	Instituto de Ingenieros Eléctricos y Electrónicos - Institute of Electrical and Electronics Engineers
IETF	Grupo de Trabajo de Ingeniería de Internet - Internet Engineering Task Force
iLBC	Internet Low Bitrate Codec
IP	Protocolo de Internet - Internet Protocol
ITU	Union Internacional de Telecomunicaciones - International Telecommunications Union
iTV	Televisión Interactiva – Interactive Television
LAN	Red de Área Local - Local Area Network
MGCP	Protocolo de Control de Pasarela de Medios - Media Gateway Control Protocol
MPEG	Grupo de Expertos en Imágenes en Movimiento - Moving Picture Expert Group
MRL	Localizador de Recursos de Medios – Media Resource Locator
MUMMS	Measuring the Usability of Multi-Media Software
NAT	Traducción de Direcciones de Red - Network Address Translation
PCM	Modulación por Pulsos Codificados – Pulse Code Modulation
PES	Flujo Elemental Paquetizado - Packetized Elementary Stream

PPV	Pago por Visión - Pay Per View
RTP	Protocolo de Transporte de Tiempo Real - Real time Transport Protocol
RTSP	Protocolo de Flujo de Tiempo Real - Real Time Streaming Protocol
RTCP	Protocolo de Control de Tiempo Real - Real Time Control Protocol
SASSI	Subjective Assessment of Speech System Interfaces
SB- ADPCM	Modulación por Pulsos Codificados Diferencial Adaptativa de sub-banda - Sub-band Adaptive Differential Pulse Code Modulation
SDP	Protocolo de Descripción de Sesión – Session Description Protocol
SIP	Protocolo de Inicio de Sesión – Session Initiation Protocol
SMPTE	Sociedad de Ingenieros de Televisión e Imágenes en Movimiento - Society of Motion Picture and Television Engineers
STB	Set Top Box
TCP	Protocolo de Control de Transmisión - Transmission Control Protocol
TS	Flujo de Transporte – Transport Stream
TTS	Texto a Voz – Text to Speech
UDP	Protocolo de Datagramas de Usuario - User Datagram Protocol
URI	Identificador de Recursos Universal - Universal Resource Identifier
URL	Localizador Uniforme de Recursos - Uniform Resource Locator
VoD	Video bajo Demanda - Video on Demand
VoIP	Voz sobre IP - Voice over IP
WiMAX	Worldwide Interoperability for Microwave Access
XMPP	eXtensible Messaging and Presence Protocol

INTRODUCCIÓN

La televisión desde su origen como servicio comercial y de entretenimiento en formato analógico, ha evolucionado hacia la televisión digital, trayendo consigo cambios en la calidad de los contenidos difundidos y la concepción en la entrega de los servicios.

Así mismo, a la televisión también se han incorporado nuevos conceptos como la interactividad, generando nuevos servicios, transformando el papel del usuario y permitiéndole realizar acciones más allá que la sintonización de un canal.

El Video bajo Demanda hace parte de los nuevos servicios asociados a la televisión, en éste se observa el cambio de rol del televidente, a través de la conducción del proceso de interacción y la ejecución de comandos simples.

A pesar de la innovación en los contenidos y los servicios de la televisión, aún se conserva el control remoto como mecanismo de entrega de mandos a distancia. De esta manera, surgen inconvenientes relacionados con la eficiencia del mismo, durante la elección y consumo de un servicio interactivo.

De acuerdo a lo anterior, es necesario indagar acerca de nuevas formas de interacción con la televisión, que permitan el acceso a los nuevos servicios de manera oportuna y simple. Por tanto, dentro del presente proyecto de investigación, se desea explorar acerca de una propuesta de interactividad basada en voz para el manejo del servicio de Video bajo Demanda.

Este documento, se encuentra organizado de la siguiente manera:

En el primer capítulo, se abordan los temas relacionados con la televisión interactiva, el cambio en el rol del usuario y la entrega de contenidos. Posteriormente se analizan las principales interfaces para la entrega de comandos a distancia, con el objetivo de identificar las ventajas y desventajas de cada una de ellas, y a partir de ello, realizar una clasificación de las mismas.

El segundo capítulo, describe los conceptos y elementos esenciales de un escenario básico para el suministro de servicio de Video bajo Demanda, los cuales se requiere tener

presentes para poder desarrollar un propuesta de interactividad basada en voz lo más cercana posible a un contexto real.

El tercer capítulo, contiene una descripción detallada de la solución piloto de interactividad basada en voz a usar en el proceso de evaluación de la propuesta; a esta se llega luego de un proceso de desarrollo incremental que considera algunos criterios que deben ser cumplidos para tener una solución adecuada. Dichos criterios se establecen con base en las características de las interfaces analizadas en el capítulo 1. Adicionalmente, como parte del proceso de implementación, los conceptos de un escenario de servicio de VoD, son aplicados en la configuración y puesta a punto del mismo.

El capítulo cuatro, analiza la evaluación realizada de la solución piloto para interactividad basada en voz propuesta. Esta se realiza desde dos perspectivas: la subjetiva (determinar la percepción del usuario al usar la interfaz basada en voz) y la objetiva (medidas de parámetros de red cuando se ejecutan los comandos de voz).

Finalmente, en el capítulo cinco se presentan las conclusiones, recomendaciones y trabajos futuros, resultado del desarrollo del presente proyecto.

CAPÍTULO 1. LA TELEVISIÓN Y LOS MECANISMOS DE INTERACTIVIDAD

1.1 INTRODUCCIÓN

La televisión es uno de los medios de comunicación y entretenimiento de mayor uso por parte de los seres humanos, se encuentra en constante evolución, incrementando la calidad de los contenidos emitidos y presentando mejoras que benefician al televidente.

La forma tradicional en que los usuarios eligen dichos programas e interactúan con la televisión ha sido el control remoto, un dispositivo integrado a dicho ambiente desde hace varios años, lo que ha acostumbrado al televidente a su uso. Los trabajos [1], [2], [3], [4] han reportado que existen problemas en el diseño del mismo, como etiquetas ambiguas y en otros, demasiados botones, entre otros, indicando que estos deberían ser solucionados.

Junto al control remoto existen otros mecanismos de interacción que pueden ser considerados como alternativas de uso por parte del usuario final al momento de utilizar sus servicios de entretenimiento, entre estos están: la interfaz basada en voz, las interfaces táctiles, las interfaces gestuales y las interfaces multimodales, cada uno de ellos buscando mejorar el proceso de interacción, de tal forma que pueda ser más flexible y sencillo.

Dentro de este capítulo, se describe la evolución del concepto de la televisión tradicional hacia la televisión interactiva. Además, se identifican las características, ventajas y desventajas de algunos mecanismos de interacción, permitiendo realizar una clasificación no excluyente ni definitiva de los mismos.

1.2 LA TELEVISIÓN

En esta sección, se presentan los principales conceptos relacionados con la evolución de la televisión, con el objetivo de conocer el contexto en el cual se utilizan los principales mecanismos de interacción.

1.2.1 Definición

De acuerdo a [5], la televisión se entiende como el contacto personal con las comunicaciones audiovisuales, considerando que la emisión es estandarizada, dependiente del tiempo, unidireccional y pública.

De la definición se destaca, que la emisión de contenidos se realiza de manera unidireccional, como consecuencia de ello, el papel del televidente es sencillamente sintonizar un canal y observarlo sin intervenir en el proceso de difusión [6]-[9].

Por otro lado, se encuentra que dichos contenidos son dependientes del tiempo, de esta manera, el televidente es un espectador, supeditado a la difusión de contenidos con horarios preestablecidos por el proveedor de servicio [6], [8], [9].

La emisión estandarizada, desencadena el envío de información común hacia todos los receptores, por consiguiente, nuevamente es el proveedor de servicio quien decide cuales son los contenidos disponibles y que observará el usuario [8].

Con el proceso de evolución hacia la televisión digital, se transforma la concepción de la emisión de los contenidos y el papel que desempeña el televidente [9].

En relación con los contenidos, algunos beneficios yacen en la reducción del ancho de banda ocupado para su transmisión, el incremento de la calidad de la imagen y el sonido, así como la oferta de programación [10], [11], [12].

Gracias a la integración de la interactividad a la televisión, el rol convencional del televidente evoluciona, permitiéndole acceder a un nuevo paradigma de emisión de contenidos, en el cual, él es el conductor del proceso [6], [9]. De esta manera, se generan nuevos servicios como: Video bajo Demanda (VoD, *Video on Demand*), Pago por Visión (PPV, *Pay per View*), tele-educación, juegos interactivos, entre otros [6], [8].

1.2.2 Interactividad

La interactividad hace parte de la comunicación en general y se consideran dos tipos: una comunicativa, en la cual los usuarios pueden enviar, recibir información o comunicarse entre ellos, y otra selectiva, que permite elegir contenidos; ambas dependientes del grado de interactividad, presente en el medio de comunicación y/o del servicio y del contexto de uso [13] [14].

Desde el punto de vista de la psicología, la interactividad emerge del concepto de interacción, que implica intercambio o influencia mutua entre participantes. De acuerdo a las comunicaciones, se refiere a las acciones de los televidentes frente a un contenido, y

determina la habilidad potencial de un medio de comunicación para permitirle al usuario, influenciar el contenido y/o la comunicación [14].

Para que exista interactividad, es necesaria la participación de mínimo dos partes (Persona - Persona, Persona – Computador, Persona – Televisor, entre otros.), los cuales deben compartir un entorno común, que sirva como medio de intercambio de información; donde los roles de emisor y el receptor pueden ser intercambiados, y estos a su vez, pueden modificar los mensajes compartidos [15].

En concordancia con [16], la interactividad es un proceso dinámico e interdependiente entre un emisor y un receptor, que depende de las posibilidades que ofrece el medio tecnológico usado para el intercambio de mensajes entre los mismos. Desde otro punto de vista, la interactividad es la habilidad potencial del medio, para permitirle al usuario influenciar el contenido y/o la forma de comunicación mediada.

En relación con la televisión, se considera que a través de la interactividad los proveedores de contenidos pueden obtener información del perfil del televidente, conllevando a una entrega de contenidos y publicidad personalizada, ayudando así a mejorar la satisfacción del cliente con el servicio ofrecido [17] - [19].

1.2.3 Televisión Interactiva

De acuerdo a la Corporación de Radiodifusión Británica (BBC, *British Broadcasting Corporation*) [20], la Televisión Interactiva (iTV, *Interactive Television*) se define como: “los contenidos y los servicios (adicionales a la televisión lineal y los canales de radio), los cuales están disponibles para los televidentes digitales, a través de la navegación en la pantalla del televisor. En la práctica, hasta este momento, esto significa darle al usuario el control sobre el video, el audio, los gráficos y los elementos de texto, o permitirle el uso de juegos simples y cuestionarios, o enviar comunicaciones sencillas al proveedor de contenido”.

De la definición de iTV, se destaca el control que se le brinda al usuario sobre los contenidos y los servicios, así como, la comunicación que se puede establecer con el proveedor de contenido [6], [14]. Además, es importante que la definición de iTV no se encuentre ligada a un prototipo de servicio, un dispositivo multimedia en particular, ó a ciertas características que se deben cumplir, evitando el riesgo de incluir servicios que no son interactivos y excluir algunos que sí lo son [14].

Los beneficios de la interactividad se pueden resumir en control, personalización y selección del contenido a consumir a través de pantallas de navegación; intercambio de información; retroalimentación hacia el proveedor de contenido a través del enlace de retorno; enriquecimiento y participación en los programas y/o servicios; y crecimiento del nicho de mercado para los proveedores de servicios y contenidos [6].

iTV trae consigo un cambio en la forma en que los contenidos se entregan [8], pasando de un modelo *Push*, en el cual el contenido entregado es estático y lineal, a un modelo *Pull*, en el cual el televidente selecciona los servicios a los cuales desea acceder, influenciando de alguna manera el contenido que observa, llegando a ser el conductor del proceso que ocurre entre él y el proveedor de contenido [6].

Acorde a la participación del usuario en el proceso de interacción, se tienen los niveles de interactividad descritos en la tabla 1 [21], los cuales demarcan las acciones que un usuario puede realizar cuando hace uso de un servicio interactivo en particular, y que se proponen en el marco de servicios para Televisión Interactiva.

Tabla 1. Niveles de Interactividad de Acuerdo a las acciones a realizar.

Nivel de Interactividad	Tipo de Acción	Aplicaciones y/o Servicios
0	Elección de Canales y Programas.	<ul style="list-style-type: none"> ▪ Sistemas de Acceso Condicional. ▪ Guía de Programación Electrónica (EPG, <i>Electronic Program Guide</i>). ▪ Televisión Tradicional.
1-3	Elección y operaciones desde menús.	<ul style="list-style-type: none"> ▪ VoD (reproducir, pausar, detener, etc.). ▪ Personalización (elección de cámara, ángulo, etc.). ▪ Canales adicionales (para información extra o para publicidad). ▪ Comercio Electrónico.
3	Producción de Información	<ul style="list-style-type: none"> ▪ Participación y Contribución en programas ▪ Producción de canales o programas por parte de los televidentes.
4	Intercambio de información y comunicación.	<ul style="list-style-type: none"> ▪ Intercambio de información y/o comunicación entre los televidentes.

Si bien los niveles de interactividad establecidos en la tabla 1, se propusieron hace varios años, las acciones relacionadas al manejo de los servicios aún es vigente, esto se verifica al analizar la clasificación propuesta por [22], la cual se observa en la tabla 2. En esta aproximación se definen 7 niveles de interactividad, los cuales están asociados al tipo de servicio con el cual se interactúa, de su análisis se identifica que los servicios responden a ciertas acciones, las cuales se indican en la tabla 3. De la tabla 2, se destaca el nivel 7, en el que se plantea que el televidente no solo envía información, si no que puede incluso modificar el contenido en sí mismo.

Tabla 2. Niveles de Interactividad de acuerdo al servicio.

Nivel de Interactividad	Nombre del Servicio	Descripción
1	Televisión básica	Hace referencia a la televisión tradicional donde el televidente solo cambia de canal.
2	<i>Call in TV</i>	No existe un canal de retorno explícito. El usuario hace uso de mensajes de texto para interactuar con el servicio.
3	Televisión en	El usuario accede a programas con características de

	Paralelo	multilinguaje o multi-cámara.
4	Televisión Adicional	Hace referencia a la televisión mejorada, en la cual el televidente tiene la opción de acceder a información relacionada con el programa que observa.
5	Servicios bajo Demanda	Hacen uso de un canal de retorno explícito para que el usuario pueda acceder a los servicios.
6	Televisión Comunicativa	Permite el acceso a contenidos multimedia desde otros medios de comunicación como el Internet. También se puede hacer uso de chats o el envío de mensajes electrónicos.
7	Televisión Completamente Interactiva	Es el más alto nivel de interactividad, en la cual el usuario puede realizar su propia historia y modificar el contenido.

Tabla 3. Niveles de Interactividad de acuerdo a las acciones y los servicios.

Nivel	Nombre	Acciones	Servicios
1	Televisión Básica	<ul style="list-style-type: none"> Elección de canales y programas 	<ul style="list-style-type: none"> Televisión tradicional
2	Call in TV	<ul style="list-style-type: none"> Participación y contribución en programas. 	<ul style="list-style-type: none"> Votaciones en programas.
3	Televisión en paralelo	<ul style="list-style-type: none"> Elección y operación desde menús: cambio de lenguajes 	<ul style="list-style-type: none"> Personalización (elección de cámara, ángulo, etc.). Canales adicionales (para información extra o para publicidad).
4	Televisión Adicional	<ul style="list-style-type: none"> Elección de canales y programas 	<ul style="list-style-type: none"> Canales adicionales (para información extra). No necesita canal de retorno.
5	Servicios bajo Demanda	<ul style="list-style-type: none"> Elección y operación desde menús: cambio de lenguajes 	<ul style="list-style-type: none"> Video bajo Demanda
6	Televisión Comunicativa	<ul style="list-style-type: none"> Intercambio de información y comunicación. 	<ul style="list-style-type: none"> Correo electrónico. Juegos en línea.
7	Televisión Completamente Interactiva	<ul style="list-style-type: none"> Modificación del contenido. 	-----

También se puede tener otra posible forma de clasificar los mecanismos de interactividad, la cual se hace basada en el lugar físico donde residen los dispositivos con los cuales se hace la ejecución de las acciones de interacción por parte de los usuarios, estos se sintetizan en la tabla 4. En la interactividad *Local* toda la información se encuentra previamente descargada en el *Set Top Box* (STB), con el cual se interactúa para la ejecución de las acciones ordenadas por el usuario, adicionalmente en esta aproximación, no se hace uso de un canal de retorno [5]. En la interactividad denominada en este trabajo *Remota*, las acciones de interactividad se realizan en los dispositivos del proveedor de servicio, y se tienen dos formas: *Simple* y *Completa*. En la interactividad *Simple*, se

requiere un canal de retorno, no obstante, la respuesta al usuario no se entrega sobre el mismo canal, si no a través de la pantalla del televisor. La interactividad *Completa*, utiliza el canal de retorno de manera bidireccional, permitiendo al televidente acceder a servicios a través de Internet, desde un computador [5], [21], [23].

Tabla 4. Tipos de interactividad dependiendo del lugar de ejecución de los comandos.

Tipo de Interactividad	Nombre	Característica	Servicios Disponibles
Local (Hogar - STB)	Local	Sin canal de retorno.	<ul style="list-style-type: none"> ▪ Portales informativos. ▪ Teletexto. ▪ Cerca video bajo demanda.
Remota (Proveedor – Servidores)	Simple	Con canal de retorno, con respuesta en pantalla.	<ul style="list-style-type: none"> ▪ Verdadero VoD. ▪ Pago por visión. ▪ Compras desde el hogar (homeshopping).
	Completa	Uso del canal de retorno de manera bidireccional.	<ul style="list-style-type: none"> ▪ Correo electrónico. ▪ Apuestas. ▪ Juegos en línea.

Al observar la Tabla 3, se destaca el aporte de la interactividad a la televisión convencional, en la cual se incrementa el número de servicios. Así mismo, se nota el papel activo del televidente, donde puede participar de programas y servicios, producir y compartir contenidos propios, incluso comunicarse con otros televidentes, como si lo hiciera a través de otros medios de comunicación como Internet [9], [14].

Ahora bien, con la evolución de la televisión y la llegada de nuevos servicios, es necesario el uso de mecanismos de interactividad adecuados a las nuevas necesidades de intercambio de información.

1.3 MECANISMOS DE INTERACTIVIDAD EN UN AMBIENTE DE DISTRIBUCIÓN DE TELEVISIÓN

En esta sección se identifican las características, ventajas y desventajas de algunos mecanismos de interactividad, y con base en ellas se realiza la clasificación de los mismos.

1.3.1 Principales mecanismos de Interacción en un ambiente de televisión

Acorde a los proyectos [1] – [3], [24] - [29], se han identificado el control remoto, la voz, las interfaces táctiles, multimodales y gestuales, como las principales formas de interactividad con la televisión.

Entre otros dispositivos utilizados para el proceso de interacción, se encuentran: un apuntador láser, los teléfonos móviles, el teclado de computador y un cubo con sensores de inclinación.

Acorde a [4], [26], [30], [31], se han establecido unas propiedades importantes, que deben ser consideradas en la caracterización de los mecanismos de interactividad. Estas son:

- ✓ Realimentación: corresponde a la entrega de respuestas al usuario cuando éste ha entregado una orden [4].
- ✓ Eficiencia: tiempo requerido para completar una tarea [30].
- ✓ Naturalidad: hace referencia al uso intuitivo de una interfaz, en el cual, el televidente intuye lo que debe hacer, sin que sea necesario entrenamiento previo [31].
- ✓ Carga Cognitiva: relación entre el conocimiento y nivel de concentración necesario para realizar una tarea [26].

A continuación, se explican los principales mecanismos de interacción, así como la relación con las características mencionadas.

1.3.1.1 El Control Remoto

El control remoto es la forma tradicional para la entrega de comandos a distancia, empleado en la interacción con la televisión y otros electrodomésticos. Por tanto, el usuario lo acepta como una extensión del cuerpo, y considera que no debe aprender acerca de otras interfaces, al estar familiarizado con su uso [26]. Así mismo, recibe realimentación tangible, por medio de la presión ejercida sobre los botones, incrementando el grado de inmersión en el ambiente de interacción [1], [32].

En relación con iTV, se considera que el control remoto no permite un proceso de interacción intuitivo y sencillo [1], [3], [26], [27] [33], [34], y es posible que para acceder a algunos servicios sean necesarios un gran número de pasos [24]. Además presenta mayor carga cognitiva, como consecuencia del número de etiquetas asociadas a los botones, las cuales se encuentran en diferentes idiomas o cambian de posición entre el diseño de un control remoto y otro [1], [2]. Finalmente, la entrada de datos y el acceso a diferentes puntos de la pantalla, son otros problemas reportados [1], [3], [28].

Como ventaja de este dispositivo, se cuenta el uso de tecnologías de bajo costo, con baja complejidad tecnológica para su puesta en funcionamiento [34], no obstante, no existe un diseño estandarizado y existen varios controles en el mismo ambiente de entretenimiento [2], [4].

1.3.1.2 La Voz

Una interfaz por voz se reconoce como natural, intuitiva, debido a que es la forma natural de comunicación entre los seres humanos [26].

Para la interacción con iTV, la voz facilita la entrada de datos, brinda acceso a un mayor número de usuarios y disminuye el número de pasos para acceder a un servicio [1], [27]. Además, ésta requiere menos esfuerzo para la entrega de comandos y no desvía la atención de la pantalla [1].

Para una interfaz por voz, es recomendable emplear formas de activación/desactivación [36], para disminuir la probabilidad de error, como consecuencia del ambiente ruidoso en el cual se interactúa, así como, la entrega de comandos por error [27], [36].

Se recomienda utilizar sistemas de ayuda que guíen al usuario, y prevengan la confusión o entrada de comandos erróneos, que afecten la percepción del usuario frente al servicio [1], [24], [29], [34]. Finalmente, la realimentación se debe entregar por medio de mensajes en pantalla, ya que, las instrucciones a través de comandos de voz perturban la visualización de los contenidos presentados en el televisor [26], [37].

El uso de palabras preestablecidas en la gramática del sistema es una restricción para una interfaz por voz independiente del usuario, razón por la cual se pueden presentar errores de reconocimiento. Por tanto, es importante evitar el empleo de palabras homófonas y tener en cuenta que éstas sean de uso común [26].

La carga cognitiva de una interfaz por voz está condicionada al número de comandos y al tiempo de interacción, puesto que, para largas sesiones el usuario puede sentir cansancio [1], [36].

La voz permite el acceso a la información de un mayor número de personas en comparación con el control remoto [3]. Además, demanda poco esfuerzo y hace que no se pierda la atención de la pantalla, porque no se mueve el cuerpo y no es necesario mirar el control u otro dispositivo al momento de entregar un comando [1]

1.3.1.3 Interfaz Táctil

La interfaz táctil permite el acceso a la información por medio de pantallas sensibles al tacto. Éstas se emplean en computadores y dispositivos móviles, sin embargo, han sido consideradas como una opción de interacción para la televisión [29].

Este tipo de interfaz (figura 1) se integra generalmente al control remoto, entregando realimentación tangible a través del tacto [32] y reduciendo el número de botones en el mismo [1]. Dicha integración facilita la aceptación de la interfaz [1] y evita la inserción en el ambiente de interacción de un nuevo dispositivo que deba ser manipulado [28], [32].

En la interacción con iTV, una interfaz táctil facilita el alcance a cualquier punto en la pantalla, reduciendo el número de iteraciones necesaria para completar una tarea [38].



Figura 1. Interfaz táctil asociada al control remoto.

Con una interfaz táctil, la carga cognitiva se reduce en comparación con los otros mecanismos de interacción, debido a que, el usuario no debe aprender comandos, y de forma natural deduce la parte de la pantalla que debe presionar [34]. Sin embargo, en algunos casos este tipo de interfaces requiere la lectura de manuales para configurarlas y utilizarlas correctamente [3].

Dada la distancia a la cual se encuentra el usuario del televisor, no se recomienda integrar la interfaz táctil a la pantalla del televisor, ya que se puede ocasionar cansancio al televidente [34].

1.3.1.4 Interfaz Gestual

Los gestos son un tipo de comunicación no verbal realizada con partes del cuerpo. Éstos han sido pensados para la interacción con la televisión [1], debido a que el reconocimiento se realiza a través de cámaras y no es necesaria la adición de un dispositivo externo que deba ser manipulado [35].

Como consecuencia del uso de cámaras, es posible que el televidente sienta la pérdida de su privacidad, por tanto, algunos sistemas centran el reconocimiento en algunas partes del cuerpo o del ambiente de entretenimiento, como las manos o mesas de centro [39].

Para hacer uso de una interfaz gestual, los usuarios deben aprender la forma de los comandos a entregar, convirtiéndola en un mecanismo de interacción poco natural e intuitivo, generando confusión, desorientación y carga cognitiva [34], [35], [40]. Además, se considera poco eficiente, debido a que para completar una tarea puede ser necesario el uso de un gran número de comandos [1].

Generalmente, este tipo de interfaces posibilitan la manipulación directa de los objetos en pantalla, permitiendo la entrega de realimentación visual al televidente durante la interacción [1], [41].

Al interactuar con la televisión a través de los gestos, se deben evitar largas sesiones, ya que pudiera causarse cansancio muscular [39]. Así mismo, no es recomendable un ambiente de entretenimiento cargado de elementos que genere problemas de reconocimiento y alta carga de procesamiento en el sistema [1], [39], [42].

La interfaz gestual, es el único mecanismo de interacción que permite la entrega de comandos por parte de varios usuarios en el mismo instante de tiempo [39].

1.3.1.5 Interfaces Multimodales

Las interfaces multimodales, emplean diversas maneras de suministrar y obtener información de un servicio, brindándole al usuario la posibilidad de elegir la forma de completar una tarea [29], [37]. Éstas incrementan la eficiencia y reducen la probabilidad de error [36], ya que las deficiencias de una interfaz son compensadas con las ventajas de otras [29].

Este tipo de mecanismos de interacción requieren altos niveles de procesamiento, sincronismo y coordinación, como consecuencia del incremento en el número de comandos a reconocer desde diferentes interfaces de entrada [29].

En relación con iTV, las interfaces multimodales permiten de manera exitosa la interacción con los nuevos servicios, como el video bajo demanda [43]. Así mismo, posibilita el acceso a la información por parte de todos los usuarios [44].

Si el televidente no ha tenido experiencia previa con las interfaces multimodales, es posible que este deba aprender por medio de ensayo y error [43], generando quizás una experiencia poco satisfactoria [26].

1.3.1.6 Otras Interfaces

Además de los mecanismos de interacción antes mencionados, se han diseñado otro tipo de interfaces, como el proyecto MagicWand [45], que utiliza un apuntador láser y una cámara para controlar la pantalla del televisor. También, existe un cubo con sensores de orientación, figura 2, que controla la pantalla tomando en cuenta la inclinación del cubo. La conexión entre el cubo y el televisor se realiza a través de infrarrojo [46].

Otra interfaz encontrada, es un lapicero, el cual permite el acceso a los contenidos, cuando lee sobre un papel en el cual han sido impresas algunas instrucciones pre-establecidas para el cambio de canal y otras opciones relacionadas con la televisión [33].



Figura 2. Interacción mediante un cubo.

Dentro de la adaptación de otras interfaces para su uso en la televisión, se encuentra un teléfono celular [47] y el teclado, que no es muy utilizado debido al sentimiento de rechazo que genera por su asociación con el computador y un ambiente de trabajo [32].

Luego del análisis de algunos mecanismos de interacción en un ambiente de televisión, la Tabla 5, presenta una síntesis de las principales características, ventajas y desventajas de los mismos.

Con la identificación y descripción de otras interfaces para la entrega de comandos en los servicios de televisión, se concluye el análisis de las características, ventajas y desventajas, de los principales mecanismos de interacción utilizados en un ambiente de televisión.

Tabla 5. Principales Mecanismos de Interacción

Tipo de Interfaz	Propiedades Características de la Interfaz				Ventajas	Desventajas
	Realimentación	Eficiencia	Natural e Intuitiva	Carga Cognitiva		
Control Remoto	Por medio del tacto cuando presiona un botón.	Entrada de datos limitada, confusión en el televidente y puede tomar más tiempo acceder a un servicio.	El usuario ya conoce los botones tradicionales.	El usuario debe buscar el botón correcto y tratar de entender las etiquetas en otros idiomas.	<ul style="list-style-type: none"> ▪ Tecnología de bajo costo, con baja complejidad tecnológica para su puesta en funcionamiento. 	<ul style="list-style-type: none"> ▪ No existe un diseño estandar de la interfaz. ▪ No facilita el acceso de todo tipo de usuario. ▪ En el mismo lugar existen varios controles.
Voz	La realimentación se debe entregar en pantalla, no a través de la voz.	Facilita la entrada de datos y disminuye el tiempo necesario para lograr acceder a un servicio.	Forma natural de comunicación de los seres humanos.	Depende del número de comandos disponibles y de la duración de la sección interactiva.	<ul style="list-style-type: none"> ▪ Oculta al usuario la complejidad del sistema. ▪ Brinda comodidad al televidente. ▪ No se desvía la atención de la pantalla. ▪ Acceso a la información por parte de la mayoría de los usuarios. 	<ul style="list-style-type: none"> ▪ Se presentan errores de reconocimiento de la voz en un ambiente ruidoso. ▪ Dependiente de la gramática del sistema. ▪ Entrega de comandos por error.
Táctil	Realimentación a través del tacto.	Permite el alcance de cualquier punto en la pantalla, reduciendo el número de iteraciones para completar una tarea.	El usuario no debe aprender de su uso por medio de manuales.	El televidente no debe aprender comandos o buscar botones.	<ul style="list-style-type: none"> ▪ No se requiere otro dispositivo a parte del control remoto. ▪ Disminución de botones en el control remoto. 	<ul style="list-style-type: none"> ▪ Se desvía la atención de la pantalla.
Gestual	Depende de la realimentación en pantalla.	Puede tomar más tiempo la realización de un comando que la presión de un botón del control remoto.	Se debe aprender la forma de los comandos.	Los comandos pueden generar confusión y desorientación.	<ul style="list-style-type: none"> ▪ Permite la interacción de varios usuarios en el mismo instante. ▪ No se requieren dispositivos externos que deban ser manipulados. 	<ul style="list-style-type: none"> ▪ Pérdida de privacidad. ▪ Altos requerimientos en hardware y software. ▪ Dependiente del ambiente. ▪ Activación por error.

1.3.2 Clasificación de los Mecanismos de Interactividad

Con base en las características, ventajas y desventajas de las interfaces antes descritas, a continuación se proponen 3 clasificaciones desde el punto de vista de las acciones a desarrollar, los usuarios y las tecnologías involucradas, estas clasificaciones no son excluyentes ni definitivas.

1.3.2.1 Por las Acciones a Desarrollar

La Tabla 6, contiene la clasificación de las interfaces de acuerdo a las acciones que se desean desempeñar. Ésta toma en cuenta las acciones relacionadas con los niveles de interactividad presentes en la Tabla 3. De las acciones se destaca la búsqueda y elección desde menús, debido a que estas acciones están relacionadas con el VoD. Las recomendaciones de la Tabla 6, se realizan tomando en cuenta el trabajo realizado en [3], [39], [40], [48].

A la tabla 6 se agrega lo que se ha denominado en el presente trabajo el “Grado de Movilidad”, refiriéndose a que tan libre es el usuario de moverse dentro del ambiente de entretenimiento mientras esta haciendo uso de la interfaz. Se considera importante el grado de movilidad porque de esta manera el usuario no se debe ubicar en una posición determinada frente al televisor y puede entregar los comandos desde cualquier lugar, generando sensación de libertad. El grado de movilidad es alto, cuando el usuario puede moverse libremente, y bajo cuando debe permanecer cerca del televisor para entregar los comandos.

Tabla 6. Clasificación según las Acciones a Desarrollar.

Mecanismo de Interactividad	Acciones			Grado de Movilidad
	Búsqueda y elección de canales	Envío de datos	Acceso a la información	
Control Remoto	Si	Si	Si	Alto
Voz	Si	Si	Si	Medio
Gestual	Si	No	No	Bajo
Táctil	Si	Si	Si	Bajo - Alto
Multimodal	Si	Si	Si	Alto

El control remoto posee un grado de movilidad alto gracias a las tecnologías inalámbricas empleadas. Para la búsqueda y elección de canales el control remoto es apropiado, dado que el televidente se encuentra familiarizado con los botones asociados a dichas acciones. El control remoto puede ser empleado para el envío de datos y el acceso a la información [3], no obstante se considera que tiene limitaciones para el desempeño dado que puede tomar mayor tiempo lograr acceder a un servicio [1].

La voz se considera como uno de los mecanismos más adecuados para interactuar con los nuevos servicios de la televisión ya que permite desarrollar las tareas de forma eficaz. Para la elección de canales, búsqueda y elección desde menús, la voz tiene es adecuada porque permite la entrada de datos de forma rápida [1]. La movilidad para la voz tiene un grado de movilidad medio, debido a que el televidente debe permanecer cerca del micrófono, sin embargo, si la interfaz por voz se asocia al control remoto el grado de movilidad puede llegar a ser alto [3].

Las interfaces gestuales son recomendadas para la búsqueda y elección de canales, porque los comandos de entrada son limitados, sin embargo no se recomiendan para las otras acciones, debido a que sería necesario incrementar nuevos comandos que lo permitan [1], [34], [40], [42]. Estas interfaces presentan bajo grado de movilidad, porque para la indicación de los comandos es necesario que el televidente esté frente a la cámara de reconocimiento [42].

Dentro del proyecto de grado se considera que el grado de movilidad de una interfaz táctil es bajo, si la interfaz se asocia a la pantalla del televisor, no obstante, si la interfaz está integrada en el control remoto la puntuación puede ser más alta. La interfaz táctil se recomienda para la búsqueda de canales, puesto que se facilita la elección de los mismos por medio de la configuración en pantalla de logos asociados a cada canal. El envío de datos y el acceso a la información, son recomendadas para ésta interfaz debido a que podría facilitarse llegar a cualquier punto de la pantalla del televisor, como si hiciera con el mouse del computador [3].

Finalmente, las interfaces multimodales son recomendadas para realizar las diferentes acciones y poseen un alto grado de movilidad por ser una combinación de las demás interfaces.

1.3.2.2 Según la Población Objetivo

La siguiente clasificación considera la población con problemas físicos y motores. Estas recomendaciones se realizan con base en [3], [34], [40], [49].

Con la intención de brindar acceso universal es importante que las personas con problemas motores y físicos cuenten con acceso a la información por medio de interfaces adecuadas. De esta manera, la Tabla 7, contiene la clasificación de los mecanismos de interactividad de acuerdo a las limitaciones físicas y motoras.

Las interfaces gestuales no son recomendadas a las personas con problemas visuales, debido a que pueden presentarse problemas en el aprendizaje y la indicación de los comandos. Así mismo, el uso del control remoto es restringido, como consecuencia del amplio número de botones incluidos en el mismo, lo que puede causar confusión en el momento de entregar el comando. Finalmente, las interfaces táctiles no son sugeridas, ya que se requiere observar la pantalla para la interacción.

Tabla 7. Clasificación de acuerdo a los usuarios con problemas físicos.

Problemas Físicos	Mecanismo de Interactividad				
	Control Remoto	Voz	Gestual	Táctil	Multimodal
Discapacidad visual	No	Si	No	No	Si
Discapacidad auditiva	Si	No	Si	Si	Si
Problemas del habla	Si	No	Si	Si	Si
Problemas motores	No	Si	No	No	Si

Las personas con discapacidades auditivas y del habla pueden hacer uso de cualquier interfaz, en la cual no se involucre la voz. Estos usuarios cuentan con la ventaja de la comunicación por señas, a la cual se encuentran habituados, facilitando el aprendizaje de los comandos necesarios en una interfaz gestual.

En cuanto a las personas con problemas motores se recomienda solo la voz, puesto que podrían presentarse inconvenientes, dada la limitación física de los usuarios, durante el manejo del control remoto, la interfaz táctil o la realización de los gestos.

1.3.2.3 Por la Tecnologías Involucradas

La Tabla 8, propone una clasificación de acuerdo a la tecnología que emplea cada interfaz, para obtener los comandos de control. La voz y los gestos se consideran tecnologías de reconocimiento, puesto que los comandos son reconocidos a través de sensores ópticos, cámaras o micrófonos sensibles al ruido. Posteriormente, éstos se analizan, procesan y comparan con patrones preestablecidos, empleando algoritmos basados en redes neuronales y cadenas ocultas de Markov [1].

Tabla 8. Clasificación de acuerdo a las tecnologías involucradas.

Tipo de Tecnología	Interfaz Multimodal	Tecnologías involucradas
Reconocimiento	Voz	Motor de reconocimiento, micrófonos sensibles al ruido, cadenas ocultas de Markov, redes neuronales, gestor de diálogo, etc.
	Gestos	Cámaras (infrarrojas, con detección de profundidad, RGB), acelerómetros, cadenas ocultas de Markov, Redes neuronales, sensores ópticos, tecnologías inalámbricas, etc.
Manipulación Directa	Control Remoto	Tecnologías inalámbricas (infrarrojo, Bluetooth, WiFi), tecnologías de tipo mecánico, etc.
	Táctil	Sensores de presión, tecnologías inalámbricas, dispositivos de despliegue, etc.

El control remoto y la interfaz táctil, se identifican dentro del proyecto como tecnologías de manipulación directa, gracias al contacto a través del sentido del tacto. Éstos hacen uso de tecnologías de tipo mecánico, generalmente pulsadores en el control remoto, y sensores de presión en las pantallas táctiles.

Las tecnologías de manipulación directa utilizan diferentes tipos de conexión inalámbrica para entregar los comandos al televisor, de las cuales depende el grado de movilidad y la distancia entre el dispositivo inalámbrico y el televisor. Los dispositivos de despliegue son las pantallas táctiles configurables y adaptables a las necesidades específicas de cada usuario.

Con la clasificación de los mecanismos de interactividad de acuerdo a la tecnología involucrada, se concluye el análisis acerca de las opciones con las que cuenta el televidente para interactuar con iTV.

Dentro de este capítulo, se han descrito las principales características asociadas a la televisión interactiva. Así mismo, algunos mecanismos de interacción se identificaron y clasificaron, con el objetivo de poder establecer dónde cada uno de ellos es más viable de utilizar y las restricciones que tienen para los usuarios.

Con la integración de la interactividad a la televisión, el papel pasivo del televidente ha evolucionado a un papel activo, permitiéndole realizar acciones más allá del cambio de canal. En el siguiente capítulo se describen los conceptos relacionados con el servicio de video bajo demanda, un servicio en el cual se pueden observar los beneficios descritos acerca de la interactividad. Así mismo, se describe un escenario básico general empleado en la entrega de contenidos multimedia bajo la modalidad de VoD.

CAPÍTULO 2. CONCEPTOS ESENCIALES DE UN ESCENARIO DE SERVICIO DE VIDEO BAJO DEMANDA

2.1 INTRODUCCIÓN

Entre los nuevos servicios asociados a la televisión digital está el VoD, quien tiene como elemento característico el control del televidente sobre los contenidos multimedia. Dicho control es posible a través de comandos simples como pausar, reproducir, detener, entre otros.

Para que el televidente pueda beneficiarse del control ofrecido por el VoD, se hace necesario el uso de una red de comunicaciones capaz de soportar el transporte bidireccional de información. Así mismo, ésta debe permitir el transporte de grandes flujos de información con la calidad requerida por un servicio de este tipo.

Por otro lado, para la entrega de un servicio de VoD, existen protocolos de red especializados en la difusión de contenidos multimedia en tiempo real, quienes deben cumplir ciertas características y restricciones para desempeñar dicha función.

A lo largo de este capítulo se exploran los conceptos asociados a un servicio de VoD, así como los elementos y protocolos involucrados en la entrega de los contenidos multimedia en un escenario de servicio bajo demanda. Finalmente, se describe el proceso de petición y entrega de los mismos.

2.2 FUNDAMENTOS DE VIDEO BAJO DEMANDA

2.2.1 Definición

Acorde a [50], el servicio de VoD se define como: “un servicio en el cual el suscriptor puede observar un contenido de video cuando lo desee. La premisa de operación es que el contenido esta almacenado en el servidor de VoD del proveedor. El suscriptor accede a la película desde una librería de contenidos la cual puede incluir un motor de búsqueda que accede a la descripción y puntuación de la misma. Los suscriptores típicamente posee la habilidad de pausar, reproducir, rebobinar y adelantar el contenido, o incluso detenerlo y continuar observándolo posteriormente”.

Según la definición anterior, se resalta que el usuario tiene control sobre los contenidos, los cuales se manipulan a través de comandos como pausar, reproducir, detener, rebobinar, adelantar, lo que se conoce como funciones de “*Trick Mode*” [50].

El acceso a los contenidos disponibles en un servicio de VoD, se realiza a través de una guía de contenidos, la cual almacena una descripción y valoración de los mismos. Además, ésta puede incorporar un motor de búsqueda, que permita encontrarlos acorde a ciertos criterios de búsqueda como: director, actor, género, entre otros [50], [51].

Los contenidos multimedia disponibles, deben tener en cuenta los derechos de reproducción y la comunicación con los sistemas de Gestión de Derechos Digitales (DRM, *Digital Right Managment*), para que los videos no sean copiados o manipulados ilegalmente [50], [51], [52].

Para que los contenidos multimedia sean transmitidos a cada usuario de forma independiente y en cualquier momento, se emplea el modelo de distribución Unicast, el cual establece un camino dedicado entre el proveedor de servicio y el cliente para la entrega de los mismos [11].

Dado que el modelo de distribución Unicast, puede llegar a ocupar grandes recursos de red en un determinado instante, los proveedores de servicio también emplean otros modelos de distribución como Broadcast¹ y Multicast² para la entrega de VoD [53].

De acuerdo al tipo de distribución que se emplee, el modo de entrega de los contenidos y las funciones de “*Trick Mode*” disponibles, se pueden encontrar varios tipos de VoD, éstos se sintetizan en la Tabla 9.

El último aspecto a considerar de acuerdo a la definición, es el almacenamiento de los contenidos, en discos duros de alta capacidad en los servidores del proveedor de servicio.

Generalmente, los contenidos multimedia se almacenan bajo un modelo centralizado, causando problemas de saturación en el sistema. Por tanto, con la finalidad de distribuir los contenidos disponibles entre varios servidores, y minimizar dichos problemas, una variedad de arquitecturas han sido propuestas [51].

En la Tabla 10, se presenta una síntesis de las principales arquitectura de organización de los servidores de contenidos, disponibles para desplegar un servicio de VoD. Así mismo, las ventajas y desventajas de cada una son descritas.

¹Broadcast: esquema de distribución punto-multipunto en el cual un transmisor envía la misma señal a todos los receptores.

²Multicast: esquema de distribución punto-multipunto en el cual un transmisor envía la misma señal a los miembros de un grupo multicast.

Tabla 9. Tipos de VoD.

Tipo de VoD	Descripción	Ventajas	Desventajas	Funciones de Trick Mode	Forma de Distribución
Casi VoD (Q-VoD, Quasi VoD)	En esta modalidad, cada grupo multicast está asociado a un contenido multimedia. El video se transmite cuando un mínimo de usuarios lo han solicitado y se han inscrito al grupo multicast correspondiente [54], [55].	Optimización de los recursos de red, porque éstos solo se utilizan cuando se necesitan [56].	El usuario debe esperar hasta que el grupo cuente con un número mínimo de participantes [56].	Pausar, reproducir, detener	Multicast
Cerca VoD (N-VoD, Near VoD)	La transmisión de un contenido multimedia, se realiza a través de varios canales de televisión, con tiempos de inicio diferidos entre uno y otro [51].	Disminución del tiempo de espera en comparación con Q-VoD [51].	El usuario está limitado a la fecha y hora de transmisión que designe el proveedor de servicio. El proveedor de servicio, debe destinar varios canales a transmitir el mismo contenido [51].	Adelantar, Rebobinar	Broadcast
Verdadero VoD (T-VoD, True VoD)	Es totalmente fiel a la definición, por tanto, permite que el usuario, busque, elija y observe los contenidos cuando lo desee [51].	Ofrece al televidente, independencia, personalización y control total sobre el contenido [54], [55], [57].	Alto consumo de recursos de red [51].	Todas	Unicast

Tabla 10. Arquitecturas de Organización de Servidores.

Arquitectura de organización de los servidores de contenidos	Descripción	Ventajas	Desventajas
Centralizada	Este tipo de arquitectura cuenta con un cluster de servidores, para almacenar los contenidos multimedia y responder las peticiones de los usuarios [56], [58], [59].	La implementación, gestión y configuración son simples [59].	Presenta alta vulnerabilidad a fallos y altos niveles de congestión en la red [56]. Debido al inmenso número de contenidos, los discos duros pueden coparse fácilmente. A largo plazo es poco escalable en cuanto a contenidos y número de usuarios [59].
Servidores Independientes	Para prestar el servicio, la población es dividida en zonas y cada una atendida por un servidor diferente. Se denomina de servidores independientes porque entre los servidores del sistema, no existe comunicación alguna [56], [59].	Esta arquitectura es fácilmente escalable, ya que, se puede instalar un servidor donde y cuando se requiera, sin que se afecte el sistema [59].	La tolerancia a fallos es alta en comparación con la arquitectura centralizada, porque si un servidor deja de funcionar, sólo la zona queda sin servicio [56]. Todos los contenidos están guardados en todos los servidores, disminuyendo la optimización de recursos del sistema. La actualización de los contenidos es engorrosa, puesto que, todos los contenidos deben ser copiados de forma individual [59].
Jerárquica de Servidores o con servidores Proxy	El servicio es ofrecido por medio de un servidor central, que almacena todos los contenidos, al que se comunican los servidores de menor jerarquía, denominados servidores proxy, y son éstos quienes atienden las peticiones de los usuarios [58].	Disminución en el número de replicas de los contenidos multimedia disponibles, debido a que un contenido multimedia se almacena dentro del servidor proxy, solo si ha sido solicitado [56], [59].	Dado que un servidor atiende un grupo de usuarios, si éste deja de funcionar, toda la zona queda sin servicio [59]. Se presentan cuellos de botella cuando se generan muchas peticiones [56].
Distribuida	Una red de servidores comunicados entre sí, sirven las peticiones de los usuarios. La petición llega a un servidor en particular y si éste no cuenta con el contenido solicitado, el cliente es direccionado hacia el servidor que puede atender la petición [59].	Altamente tolerante a fallos y fácilmente escalable. Cuenta con balanceo de carga, disminuyendo la probabilidad de denegación del servicio. Disminuye la capacidad de almacenamiento del sistema [56].	Requiere altas velocidades de comunicación y un sistema de gestión que sincronice los servidores.

Para desplegar cualquier tipo de VoD, se requiere una infraestructura de red capaz de cumplir con los requerimientos de este servicio y permitir la entrega del mismo en forma exitosa. A continuación, se explica un escenario de servicio de VoD y los elementos de red que intervienen en él.

2.3 ESCENARIO DE SERVICIO DE VOD

El escenario que se eligió como referente para analizar el servicio de VoD, fue el de arquitectura de organización de servidores centralizada, dada su simplicidad de implementación, reducido número de componentes, y a la facilidad para mapear dicha arquitectura a la infraestructura de equipos, red y recursos software disponibles en la FIET. En la Figura 3 se observan los componentes de la arquitectura de organización de servidores seleccionada: servidores, red de comunicaciones y dispositivos cliente [56], [57], [59].

A continuación se realiza una breve descripción de los componentes de la arquitectura seleccionada.

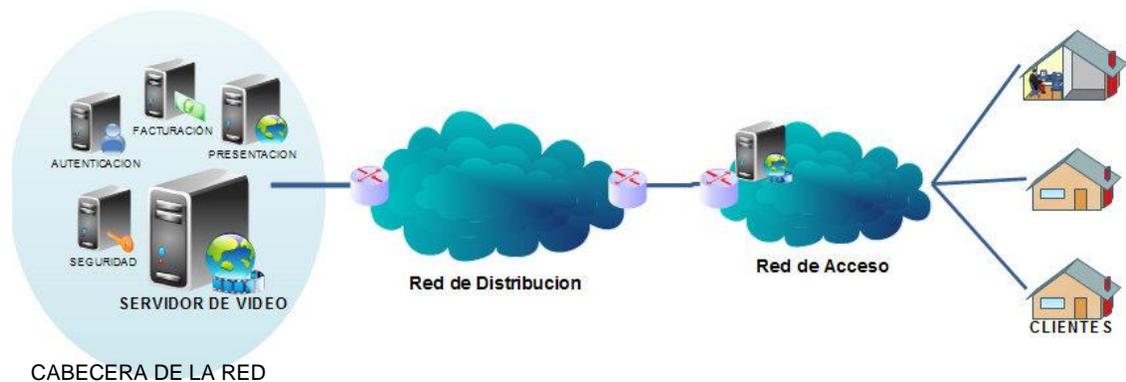


Figura 3. Escenario de Servicio de VoD.

2.3.1 Servidores

Generalmente, los servidores están ubicados en la cabecera de la red, aunque es posible encontrarlos en puntos cercanos a los usuarios dependiendo de la arquitectura.

Los servidores más importantes son los de video, puesto que son el núcleo de un servicio de VoD. Éste envía el contenido multimedia a través del streaming³ y mantiene la sesión hasta que éste se ha reproducido completamente en el dispositivo cliente. De igual manera, recibe y ejecuta las peticiones de control sobre el contenido [54], [56] - [59].

³ Streaming: proceso de envío de audio y video a través de paquetes IP que permite que un video se pueda observar tan pronto dichos paquetes arriban al dispositivo cliente, evitando la descarga total de un contenido.

Entre los servidores que complementan la prestación de un servicio de Video bajo demanda en un ambiente real, se pueden encontrar: el de aplicación, el de facturación y el de seguridad. En estricto orden, el primero se encarga de verificar que el usuario está autorizado para acceder al servicio y, garantizar o denegar el acceso al mismo, dentro de éste se encuentra la lógica de funcionamiento del servicio[60]; además, identifica el servidor de video adecuado a la solicitud recibida y lo comunica con el dispositivo cliente [50]. El segundo realiza los procesos de cobro, de forma inmediata o a través de la factura mensual [11]. El tercero protege al sistema del acceso no autorizado y previene que los videos sean editados, copiados y reproducidos de forma ilegal, gracias a que incluye sistemas DRM [11], [52]. En una implementación real es vital que dichos servidores sean redundantes.

Cuando se oferta el VoD, es fundamental el uso de procesadores e interfaces de alta velocidad que le permitan comunicarse y responder en tiempo real. Así mismo, es valioso el soporte a un gran número de dispositivos cliente y una base de datos extensa con información acerca de los contenidos [54] - [57], [59]. Además de lo anterior, algunas características adicionales a tener en cuenta son la interoperabilidad, confiabilidad, escalabilidad, integración de sistemas de monitoreo y balanceo de carga [54], [57], [59].

El punto más crítico para el servidor de video, es la capacidad de almacenamiento, puesto que los videos en formato digital ocupan un tamaño considerable. Para solucionar este problema, se han generado algoritmos de codificación y compresión, denominados códec [56]. El propósito de un buen códec, es disminuir el tamaño ocupado por un video, afectando en menor medida la calidad del audio, imagen e información asociada al mismo [52], [59].

La búsqueda de un equilibrio calidad-tamaño, ha generado el desarrollo de una variedad de códecs, cada uno con la intención de mejorar el desempeño de sus predecesores. En la Tabla 11 y 12 respectivamente, se sintetizan algunos códec de audio y video utilizados en un ambiente de distribución de VoD [50], [53].

Tabla 11. Códec de Audio en un entorno IP.

Códec	Características
MPEG-1 layer II	También conocido como MP2, se recomienda su uso para transmitir en sistemas DVB. Generalmente, se utiliza en sistemas que requieren baja complejidad de codificación y decodificación. Está especificado en ISO/IEC 11172-3
MPEG-2 AAC (<i>Advanced Audio Coding</i>)	No es compatible con MPEG-1. Mejora la calidad del audio, logrando un balance entre el uso de memoria y la complejidad de procesamiento. Se especifica en ISO/IEC 13818-7.
MPEG Surround	Agrega propiedades multicanales al estándar MPEG-1 layer II y MPEG-4. Tiene compatibilidad con la codificación en mono y estéreo.
MPEG-4	La diferencia con los demás estándares, radica en que la calidad del audio se

	aumenta con menos cantidad de bits, gracias a que se agrega cierta información que permite obtener frecuencias más altas y por ende una señal de audio más fiel a la original. Se especifica en ISO/IEC 14496-3.
AC-3	Comúnmente se le llama Dolby Digital y se encuentra especificado en ETSI TS 102 366. La mayor característica es la codificación de 1 a 5.1 canales de audio, lo que genera un sonido tipo cine.
AMR-WB+	Es un algoritmo con corrección de errores que incluye interlineado y ocultamiento de errores. Se especifica en ETSI TS 126 290.

Tabla 12. Códec de Video.

Códec	Características
MPEG-2	Extiende las especificaciones de MPEG-1. Mejora la resolución de la imagen y define varios perfiles y niveles, que indican el tamaño de la imagen y cuantas imágenes pueden ser decodificadas. Está orientado hacia la transmisión de televisión y requiere canales de transmisión de mínimo 4Mbps. Se encuentra definido en ISO/IEC 13818-2 y en ITU-T H.262.
MPEG-4 Parte 10	Surge gracias a la aparición de servicios que requieren mayor calidad de imagen, con menores velocidades de transporte que las requeridas por MPEG-2. El nivel de compresión llega incluso a la mitad. Agrega el soporte a video en 3D y DRM. Es muy utilizado en sistemas bajo demanda y en el transporte de video sobre IP. Está definido en ISO/IEC 14496.
H.264/AVC (<i>Advanced Video Codec</i>)	Es similar a MPEG-4 parte 10. Es conocido por sus potentes y eficientes técnicas de compresión. Una de las desventajas son los complejos decodificadores, lo cual se compensa con la alta calidad de la imagen a bajas velocidades de transporte. Como en MPEG-4, en H.264 también se definen diferentes niveles y perfiles, combinables entre sí.
VC-1	También conocido como Windows Media 9, fue estandarizado por la Sociedad de Ingenieros de Televisión e Imágenes en Movimiento (SMPTE, <i>Society of Motion Picture and Television Engineers</i>). Este estándar define perfiles y niveles, pero a diferencia de H.264 y MPEG-2 cada nivel está especificado para cada perfil.
AVS	Codifica el video basado en la predicción espacial y temporal. En este estándar, existe un perfil y varios niveles. La paquetización de AVS, se basa en la forma de paquetización de H.264/AVC.

Continuando con el escenario de servicio de VoD, figura 3, el próximo elemento a detallar es la red comunicaciones.

2.3.2 Red de Comunicaciones

La red de comunicaciones, tiene como propósito, la transferencia de datos desde los servidores hacia los terminales de usuario y viceversa [57], [59]. Ésta se divide en red de distribución y acceso. Para desplegar una red de comunicaciones de alta velocidad, se ofrece más de una opción, entre ellas se tiene la fibra óptica, el cable coaxial, par trenzado y sistemas inalámbricos.

En la Tabla 13, se presenta de manera concisa la información acerca de las mismas. La red de distribución, traslada grandes volúmenes de información entre los servidores y puntos de acceso ubicados cerca a los usuarios, razón que amerita medios de transmisión de alta velocidad. La red de acceso, se despliega entre la red de distribución y los clientes, y aunque requiere velocidades más bajas, éstas deben contar con un mínimo ancho de banda que permita entregar los flujos de video en menos tiempo y con la menor pérdida de información posible [57] - [59].

Es importante tener en cuenta que para elegir el medio de transmisión y tecnología de transporte adecuados al servicio, se deben tener presentes criterios como la calidad de la imagen, velocidad soportada, área de cobertura, tiempos de retardo, entre otros [55] - [58].

La elección de un buen códec de video es vital, para la red de comunicaciones, debido a que entre mayor sea el grado de compresión de los mismos, menores capacidades serán necesarias para el transporte de los contenidos multimedia. Además, será posible el envío de mayor información sobre el mismo medio [52], [59].

Para la entrega de un contenido multimedia es necesario el uso de protocolos que permitan el transporte de la información de manera adecuada a través de la red de comunicaciones. A continuación se explican los protocolos necesarios para la entrega extremo a extremo de un contenido multimedia.

Tabla 13. Medios de Transmisión usados en la red de distribución y acceso.

Medio de Transmisión	Descripción	Ventajas	Desventajas	Red
Fibra Óptica	Es el medio de transmisión por excelencia en la red de distribución, donde, al emplear diferentes señales ópticas de baja longitud de onda sobre la que se envían grandes flujos de información [51], [53].	Altas velocidades de transporte [12], [51], [53]. La señal es casi inmune al ruido y a las interferencias electromagnéticas [51].	Los costos de despliegue, hacen que el proveedor restrinja su uso a la red de distribución, sin embargo lo ideal sería que se despliegue también en la red de acceso [53].	Distribución – Acceso
Cable Coaxial	Generalmente, se emplea para entregar televisión, por que permite el envío de cientos de canales, en formato analógico y digital [51], [53].	Altas velocidades. Fácil despliegue y baja complejidad en la ampliación de la cobertura [51], [53].	Con el cable coaxial, solo es posible brindar cobertura en zonas urbanas, puesto que, para zonas rurales, los costos se incrementan [51].	Acceso
Par Trenzado	Inicialmente utilizado para la transmisión de voz, las señales que viajan sobre este medio han sido modificadas, para habilitar el flujo de información a altas velocidades. Normalmente, se asocia a la tecnología de la capa de enlace conocida como Línea de Abonado Digital (DSL, <i>Digital Subscriber Line</i>) [11], [51], [53].	Existe par trenzado actualmente tendido, lo que disminuye la inversión de los proveedores de red [51], [53].	La velocidad de transporte es relativamente baja en comparación con los otros medios de transmisión. La relación entre la velocidad y la distancia es inversamente proporcional, restringiendo la cobertura alcanzable [51], [53].	Acceso
Inalámbricas	Entre las tecnologías inalámbricas consideradas para el despliegue de este servicio, se encuentra WiMAX, un estándar inalámbrico, conocido normalmente como 802.16, que para transmitir la señal, utiliza frecuencias altas en bandas licenciadas y no licenciadas [51].	Bajos costos de instalación y mantenimiento. Existen organizaciones de regulación y proveedores, encargadas de estandarizar los equipos empleados, lo que facilita la interoperabilidad de los mismos [51].	La señal sufre altos niveles de atenuación a medida que aumenta la distancia de cobertura. Presenta problemas de desempeño, causados por el clima y la interferencia de otros sistemas. Para la transmisión de contenidos multimedia no es recomendable, ya que presenta una tasa de pérdida de paquetes considerable [51].	Acceso

2.3.3 Protocolos de Red

En la Figura 4, se muestra la pila de protocolos para la entrega de un servicio de VoD, en un ambiente IP, tanto para el equipo servidor como para el terminal de usuario [51]. A continuación se describen las capas involucradas en la red de comunicaciones y sus protocolos asociados.

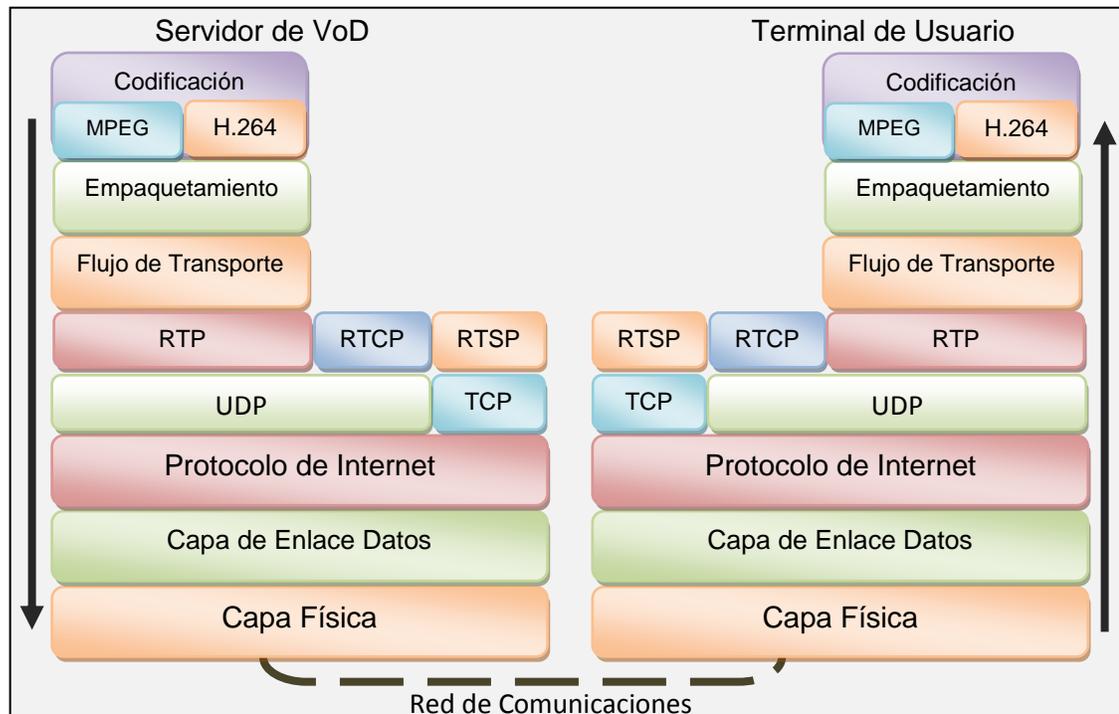


Figura 4. Pila de Protocolos.

2.3.3.1 Capa de Codificación

Se encarga de comprimir los videos almacenados en el servidor de video streaming. Utiliza los códec MPEG, H.264, entre otros. Si se utiliza H.264 como códec, ésta capa se divide en dos subcapas, una de Codificación que se encarga de comprimir el video, y una capa de abstracción de red que organiza el video en un número de paquetes. La señal de salida de esta capa es una señal digital continua en tiempo real formada por flujos elementales MPEG [51].

2.3.3.2 Capa de Empaquetamiento

Convierte los flujos elementales de video, audio y datos, en Flujo Elemental Paquetizado (PES, *Packetized Elementary Stream*), para su transmisión a través de la red digital [51], [61].

2.3.3.3 Capa de Construcción del Flujo de transporte

La función de esta capa es la construcción de un flujo de transporte continuo, el cual se logra a partir de la conversión de los PES en paquetes de Flujo de Transporte (TS, *Transport Stream*) de tamaño fijo de 188 bytes. Los TS llevan en su carga útil, video, audio o datos, pero no una combinación de ellos [51]. Entre algunos tipos de datos que pueden ser empaquetados directamente en TS, se encuentran los paquetes MPEG-1 (audio y video) y MPEG-2 (audio y video).

2.3.3.4 Protocolo de Transporte en Tiempo Real

Tal como su nombre lo indica, la función del Protocolo de Transporte de Tiempo Real (RTP, *Real time Transport Protocol*), es la transferencia de información multimedia en tiempo real, especialmente cuando se maneja streaming. Consecuentemente, cuando un servidor decide enviar un video, éste se codifica, se fragmenta en sub-paquetes, y se encapsula en la carga útil de un paquete RTP [51], [53].

Con la intención de incrementar la eficiencia en el uso del ancho de banda, la carga útil de un paquete RTP, normalmente transporta varios fragmentos de video, audio o una combinación de ambos, así como información de sincronización entre ellos [51].

Adicional a la carga útil, un paquete RTP cuenta con la cabecera del mismo, quien colabora con la identificación de los paquetes perdidos y la sincronización entre los diferentes elementos comunicados [51], [53]. Este protocolo, está diseñado para trabajar en ambientes donde el retardo y la latencia son más importantes que la pérdida de paquetes [56].

2.3.3.5 Protocolo de Flujo de Tiempo Real

La función del Protocolo de Flujo de Tiempo Real (RTSP, *Real Time Streaming Protocol*), es el transporte de las funciones de Trick Mode, que permiten controlar la sesión establecida entre el terminal de usuario y el servidor de video streaming. RTP y RTSP, funcionan para cualquier tipo de VoD, pues operan tanto para los modelos de distribución Multicast, como Unicast [51], [53], [56].

A través del Identificador de Recursos Universal (URI, *Universal Resource Identifier*), que viaja dentro de un mensaje RTSP, el servidor de presentación puede localizar la ubicación del contenido multimedia solicitado. El formato de dirección del URI, es similar al formato de una dirección URL (Uniform Resource Locator), de una página web [51], [53].

2.3.3.6 Protocolo de Control de Tiempo Real

Para conocer del estado de la sesión entre la fuente y el destino, además de información sobre la red de transporte, se ha creado el Protocolo de Control de Tiempo Real (RTCP,

Real Time Control Protocol). En el momento en que se descubre un problema, éste se corrige modificando el flujo de información y la codificación [51], [53], [56].

2.3.3.7 Protocolo de Control de Transmisión

Por ser orientado a la conexión, el Protocolo de Control de Transmisión (TCP, *Transmission Control Protocol*), garantiza la entrega de los segmentos, dado que si alguno se pierde, éste se retransmite. Como consecuencia, se generan problemas de latencia [51], [53].

El efecto que trae consigo el control automático de la tasa de transferencia de TCP, es la variación de la velocidad a la cual se transmiten los contenidos, lo que es inconveniente para la entrega de servicios en tiempo real. Por tanto, en un escenario de VoD, la función de este protocolo es transportar los mensajes RTSP [51], [53].

2.3.3.8 Protocolo de Datagramas de Usuario

Con el Protocolo de Datagramas de Usuario (UDP, *UserDatagramProtocol*), no se garantiza la llegada de los mensajes al destino, pues, éste es un protocolo no orientado a la conexión. Sin embargo, ésta es una característica ideal para el VoD, ya que los videos sufren menos retardo y latencia durante su difusión [51], [53].

Tanto TCP como UDP, cuentan con dos números de puerto en la cabecera de los segmentos y los datagramas, respectivamente, quienes facilitan identificar el equipo origen y destino. Aunque UDP no cuenta con mecanismos de re-transmisión, si posee en su cabecera un campo para la comprobación de errores, para proteger la integridad de los datos [51], [53].

Durante una sesión de VoD, los mensajes RTP y RTCP, se transportan sobre UDP. Los protocolos explicados hasta el momento, hacen parte de la capa de transporte del modelo IP.

2.3.3.9 Protocolo de Descripción de Sesión

No es suficiente el transporte y control de los contenidos en un servicio de VoD, por tanto, es necesario un protocolo que describa las características de los dispositivos y facilite la negociación de los parámetros de transmisión previo al establecimiento de la sesión. De esta manera, el Protocolo de Descripción de Sesión (SDP, *Session Description Protocol*), colabora con RTSP y entrega información acerca del tipo de contenido multimedia, el códec empleado, la sesión y el protocolo de transporte [51], [53].

2.3.3.10 Protocolo de Internet

Finalmente, en la lista de protocolos necesarios para el servicio de VoD, está un estándar de comunicaciones de red diseñado por la IETF (*Internet Engineering Task Force*), denominado protocolo de Internet (IP, *Internet Protocol*) [53].

Como resultado de la filosofía de distribución no orientada a la conexión de IP, éste puede transportar contenidos multimedia generados en tiempo real, o almacenados previamente. Así mismo, éstos se difunden a través de Unicast o Multicast [53].

Para que la carga útil de un datagrama IP alcance su destino, la cabecera de éste dispone de una dirección IP de origen y destino. Dichas direcciones, son números únicos, se asignan internacionalmente e indican la ubicación de la red y del terminal al cual debe llegar la información [51], [53].

Además de lo anterior, la cabecera cuenta con un campo de identificación, de tiempo de vida, de comprobación de errores y de tipo de servicio. En estricto orden, el primero sirve para combinar los datagramas que han sido fragmentados previamente. El segundo, se reduce en uno cada vez que atraviesa un dispositivo de enrutamiento, para que cuando llegue a cero el mensaje sea descartado. El tercero, sirve para confirmar que los datagramas no han sufrido daño alguno, pues de lo contrario deben ser descartados. Finalmente, el campo de tipo de servicio, permite que los datagramas sean transmitidos en la red con ciertos niveles de calidad de acuerdo a la carga útil que transportan [51], [53].

2.3.3.11 Capa de Enlace

Comunica la capa IP y la capa física, haciendo que esta última parezca una línea de transmisión sin errores, gracias a los sistemas de corrección de errores. Si el medio de transmisión es compartido por varios clientes, esta capa debe controlar el acceso al canal [51].

2.3.3.12 Capa Física

Es el medio de transmisión a través del cual fluye la información, entre los equipos terminales [51].

2.3.4 Terminal de Usuario

El último elemento del escenario de servicio a considerar es el terminal de usuario. Una vez éste recibe la información de la red de comunicaciones, éste la procesa y la entrega al usuario de forma que éste pueda entenderla.

2.3.4.1 Set Top Box

La integración de un procesador y una memoria de alta capacidad, hacen del STB un computador dedicado a decodificar las señales de video, para que las imágenes sean presentadas de forma adecuada. De esta manera, el STB se convierte en una interfaz entre la red y el televisor [11], [53], [56].

Para prestar los nuevos servicios interactivos, como VoD, el STB debe contar con interfaces de alta velocidad para soportar comunicaciones bidireccionales, es decir, enviar peticiones a los servidores y recibir la respuesta de los mismos. También, es importante la inclusión de discos duros, para la grabación de contenidos. Así como, buffers que permitan contrarrestar los problemas de retardo y latencia provocados por la red [11], [53], [56].

En relación con el televisor, es conveniente que el STB soporte la decodificación de imágenes en Alta Definición (HD, *High Definition*) y en formato estándar [11].

Además de lo anterior, cuando se elige un STB, es conveniente analizar los códec soportados por el mismo, para encontrar un equilibrio entre la velocidad de transporte, la calidad de la imagen y el tiempo de procesamiento [51].

Desde otra perspectiva, el tiempo de instalación, el tamaño y el costo de este dispositivo, son vitales, porque se busca que éstos sean tan bajos como sea posible [11].

Finalmente, la interacción con el STB se realiza mediante un control remoto y, opcionalmente un teclado inalámbrico [51].

2.3.4.2 Centro Multimedia

Otro terminal de usuario disponible en el mercado, es el centro multimedia, éste integra al computador, sonido de alta definición y una pantalla que permite mirar televisión, navegar en Internet, grabar contenidos y desempeñar funciones relacionadas con la edición de los mismos [11], [51].

El software integrado a este dispositivo, es un sistema operativo con la capacidad de recibir contenido bajo demanda, decodificarlo y mostrarlo en pantalla [11].

Al igual que el STB, cuenta con un control remoto y un teclado para la interacción [51].

Luego de conocer en las funciones y requisitos de los elementos pertenecientes a una arquitectura funcional para prestar el servicio de VoD, se explica el proceso de petición y entrega de un video, para de esta forma concluir la exploración de un escenario con condiciones reales.

2.3.5 Proceso de Petición y Entrega

La Figura 5, describe el proceso general de petición y entrega de un video, para una arquitectura de organización de servidores centralizada. Éste puede variar dependiendo del tipo de VoD ofertado y la arquitectura de organización de servidores empleada [50]. Algunos servidores pueden estar integrados en la misma máquina, aunque lo más recomendable es que cada uno sea independiente, para evitar los cuellos de botella [54].

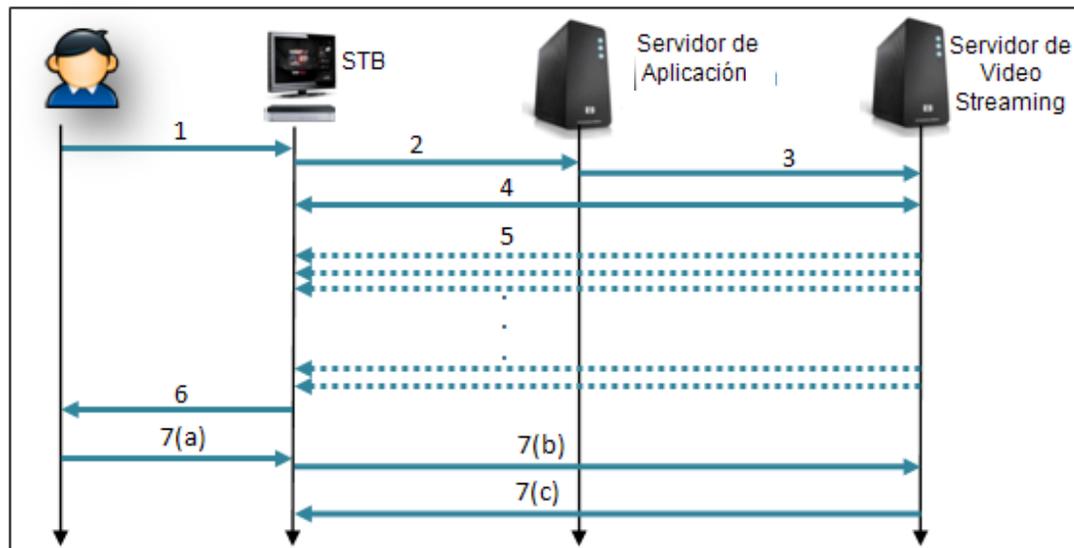


Figura 5. Proceso de Petición y Entrega.

El proceso explicado en esta sección no tiene en cuenta el servidor de seguridad, pues éste interviene generalmente en el proceso de pre-producción.

1. Búsqueda y selección del video: a través de la Guía de Programación Electrónica (EPG, *Electronic Program Guide*) almacenada en el STB, el usuario accede a la lista de contenidos disponibles y selecciona el de su interés.
2. Envío de la petición: una vez el video ha sido elegido, el STB traduce dicha petición a un mensaje que los servidores puedan entender.
3. Disponibilidad del video: el servidor de presentación verifica si el contenido puede entregarse, si es así, éste localiza el servidor de video streaming adecuado y lo comunica con el terminal de usuario. Si el video no está disponible, se envía una respuesta negativa al STB.

NOTA: Previo al envío del video, es posible que el proveedor de servicio lleve a cabo el proceso de autenticación y facturación, para agregar a la cuenta del televidente el consumo del contenido multimedia.

4. Negociación de los parámetros de transmisión: durante una corta conversación el servidor de video streaming y el STB, establecen la forma en que el video será entregado.

NOTA: Si el STB no cuenta con la licencia necesaria para la decodificación de los contenidos, éste la solicita al servidor de seguridad durante el proceso de negociación.

5. Envío del contenido multimedia: el servidor de video streaming envía los fragmentos de video, mientras estos se almacenan en el buffer del STB para su posterior decodificación y reproducción
6. Reproducción: finalmente el contenido solicitado se muestra en pantalla.
7. Si en algún momento durante la reproducción del video, (a) el televidente desea adelantar, pausar o ejecutar otra función de control, (b) el STB envía la orden para que el servidor de video streaming la ejecute, (c) y la responda.

Dentro de este capítulo, se exploraron los conceptos básicos generales acerca de un servicio de VoD, los cuales permitieron conocer que para la prestación de dicho servicio son necesarios minimamente un servidor de video, un servidor de aplicación y el cliente multimedia, los cuales se intercomunican gracias a la red de comunicaciones.

Posteriormente, la pila de protocolos involucrados en la entrega de un contenido multimedia, difundido bajo demanda, en un entorno IP fueron descritos de manera rápida. Gracias a dicha implementación, se vislumbró que los protocolos necesarios para la difusión y control de los contenidos en tiempo real, son necesarios los protocolos RTP, RTCP y RTSP.

El capítulo concluye, con la explicación del proceso de petición y entrega básico de un contenido multimedia, con el cual, se considera que los conocimientos esenciales acerca de un servicio de VoD, se han adquirido y pueden ser aplicados en el siguiente capítulo, donde se presenta la propuesta de una solución de interactividad basada en voz para el servicio de VoD.

CAPÍTULO 3. PROPUESTA DE UNA SOLUCIÓN PILOTO DE INTERACTIVIDAD BASADA EN VOZ SOBRE IP PARA EL MANEJO DEL SERVICIO DE VIDEO BAJO DEMANDA

3.1 INTRODUCCIÓN

Uno de los servicios asociados a iTV que los proveedores han seleccionado para iniciar a ofrecer interactividad es el VoD, dado que con éste, el televidente cuenta con el control sobre el proceso de interacción.

Sin embargo, se conserva el control para la entrega de comandos, y éste, dependiendo del contexto y las condiciones del servicio puede llegar a ser poco adecuado, porque tiene restringida la entrada de datos para la búsqueda de contenidos y puede tomar más tiempo navegar a través de la información sobre los videos disponibles.

Con el objetivo de brindarle al usuario nuevas formas de interacción, en las cuales éste pueda aprovechar las ventajas de los nuevos servicios y realizar el proceso interactivo de manera sencilla, dentro de este capítulo, se describe la solución piloto de interactividad basada en Voz sobre IP (VoIP, *Voice over IP*) que se propone para el control de un servicio de VoD.

3.2 PROCESO DE DESARROLLO

Para el desarrollo de la solución piloto y la integración de los componentes software que la conforman, se basó en la metodología sugerida en el Modelo de Construcción de Soluciones [62], del cual se han tomado las siguientes fases: establecimiento de los requisitos de la solución piloto, generación y descripción general de la solución piloto, Implementación, configuración y puesta a punto y evaluación (Capítulo 4), figura 6. Para este proyecto, la etapa de implementación, configuración y puesta a punto, se desarrolló por medio de actividades incrementales, incluyendo actividades de integración y verificación del correcto funcionamiento de los diferentes elementos.

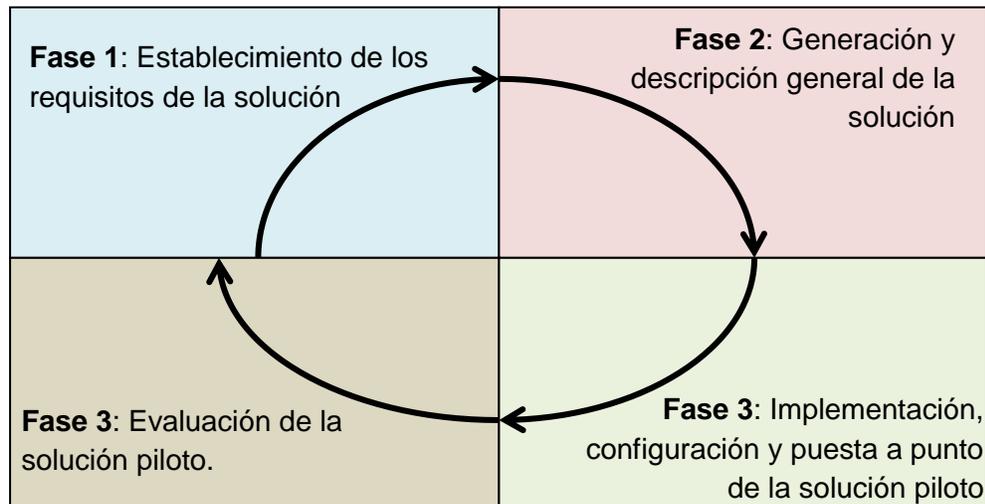


Figura 6. Fases de desarrollo del proyecto.

Para la fase de implementación, configuración y puesta a punto, se plantearon las siguientes actividades de manera incremental, figura 7:

1. **Actividad 1: Configuración de los Servidores de Video y Web:** en esta primera actividad, se decidirá entre las opciones disponibles en el mercado, cuál será el servidor de video *Streaming* y el servidor web a instalar y configurar. También se definirá la forma en la cual se integrarán los servidores antes instalados. Así mismo, se elegirá el códec bajo el cual se distribuirán los contenidos multimedia. El último paso de esta actividad es la verificación del correcto funcionamiento de los servidores.
2. **Actividad 2: Instalación y Configuración del Servidor de Comunicaciones IP y el Motor de Reconocimiento:** esta es la siguiente actividad, en la cual se elige, instala y configura el servidor de comunicaciones IP y el motor de reconocimiento. Así mismo, se define la gramática a utilizar, la cual contendrá las palabras que la solución está en condición de entender y ejecutar. Dado que es necesario un códec de audio para el transporte de la voz bajo el protocolo IP, dentro de esta fase, también se elegirá dicho códec. Esta actividad concluye con la verificación del correcto funcionamiento de este módulo.
3. **Actividad 3: Despliegue del Cliente e Integración del Sistema:** Dentro de esta última actividad se instala y configura el cliente VoIP. También se diseña e implementa la interfaz gráfica de usuario, a través de la cual éste accederá a los contenidos multimedia. Una vez configurados el servidor de comunicaciones y la interfaz, el siguiente paso será la implementación de la interfaz entre el cliente multimedia y el módulo de reconocimiento del habla, el cual permitirá la integración de la solución en su totalidad. Esta actividad, la cual concluye la fase de

instalación, configuración y puesta a punto culmina con la verificación del correcto funcionamiento de los diferentes elementos y de la solución final.

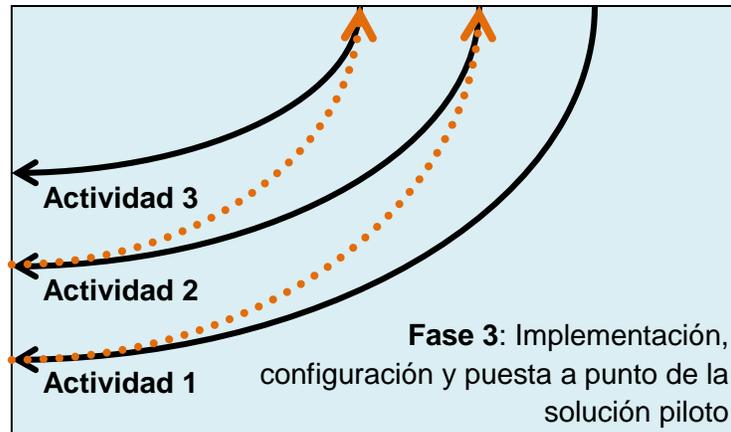


Figura 7. Actividades de la Fase de Implementación, Configuración y Puesta a Punto.

3.3 ESTABLECIMIENTO DE LOS REQUISITOS DE LA SOLUCIÓN PILOTO

Dentro de esta fase, ver figura 6 sección 3.2, se definen los requerimientos que la solución piloto seguirá. Luego de un cuidadoso estudio de los mecanismos de interacción con la televisión, el escenario de servicio de VoD, y las recomendaciones de [26] para sistemas controlados por voz en un ambiente de iTV, se encuentra que los requisitos más relevantes y que la solución debe cumplir son:

1. **Modularidad:** con el fin de facilitar la integración con otros sistemas y permitir la sustitución de las herramientas de forma rápida y sencilla, los elementos que hacen parte de la solución deben estar débilmente acoplados.
2. **Arquitectura Centralizada:** tal como se explicó en el capítulo 2, existen varias arquitecturas de organización de servidores que sirven para prestar el servicio de VoD. Para el diseño de la solución se elige una arquitectura centralizada, debido a que ésta cuenta con un servidor que atiende todas las peticiones, lo que facilita la configuración del mismo y disminuye el tiempo de puesta en funcionamiento, dado que no se debe buscar como comunicarlo con otros servidores.
3. **Motor de Reconocimiento Independiente del usuario:** para el reconocimiento del habla cuando se hace uso de motores de reconocimiento, existe la posibilidad de que estos sean con entrenamiento previo (dependiente del usuario) o sin entrenamiento previo (independiente del usuario). Para la solución se debe elegir un motor de reconocimiento independiente del usuario, para que éste haga uso directamente del servicio, sin necesidad de emplear tiempo configurando la solución propuesta.
4. **Ayuda al usuario:** para evitar tanto como sea posible generar confusión en el usuario, evitar el aprendizaje de comandos y ayudar al televidente durante la interacción con el

servicio, el sistema debe contar con una interfaz gráfica simple que indique los comandos de control, los cuales deben ser pocos y cortos. Además, la interfaz gráfica del servicio debe contar con algún tipo de ayuda, que el usuario pueda consultar cuando lo desee.

5. **Guía de contenidos:** Es necesario desarrollar una guía de contenidos que permita a los usuarios tener una percepción cercana a la que suministran los sistemas comerciales, facilitando el acceso a los contenidos multimedia disponibles y a su información asociada.
6. **Realimentación:** el sistema debe brindar realimentación a través de la pantalla, para que el televidente conozca en que parte del proceso se encuentra y la orden que ha entregado.
7. **Activación/Desactivación:** es necesario que la solución cuente con un sistema de activación/desactivación de la interfaz por voz, para disminuir la probabilidad de entrega de comandos erróneos durante los intervalos de tiempo en los cuales el televidente no está interactuando con la solución o se encuentra en un ambiente ruidoso.
8. **Bajo tiempo de respuesta:** es importante contar con herramientas que disminuyan tanto como sea posible el tiempo de procesamiento, para evitar afectar de manera negativa la percepción del usuario.
9. **Protocolos estandarizados:** con el uso de protocolos estandarizados y de amplia utilización, los usuarios podrán acceder al servicio de VoD desde cualquier terminal de usuario, sin tener que comprar un dispositivo nuevo con soporte a protocolos específicos.

Tomando en cuenta lo recomendado en [26], se han integrado los anteriores requerimientos dentro de la construcción de la propuesta, siendo posible generar una solución piloto para probar la interactividad basada en voz sobre un servicio de VoD. A continuación se explica una solución piloto de interactividad basada en voz, orientada a un ambiente IP.

3.4 GENERACIÓN Y DESCRIPCIÓN GENERAL DE LA SOLUCIÓN

Dentro de esta sección y fase del proyecto, ver figura 6 sección 3.2, se realiza un esquema general de la solución piloto a desarrollar con base en los requisitos y funciones que ésta debe cumplir, explicados en la sección anterior.

Analizando los elementos utilizados en el escenario de un servicio de VoD, y el proceso de petición y entrega de un contenido multimedia [50], se encuentra que para el servicio de VoD en un ambiente IP, los elementos software básicos requeridos son:

- ✓ **Servidor de Video:** la función de este es transmitir los flujos de video a través de streaming, para que éstos puedan ser reproducidos en el terminal de usuario.
- ✓ **Servidor de Aplicación:** este servidor es necesario porque es quien almacena la guía de contenidos, consulta si el televidente puede acceder al servicio y realiza la interconexión entre el servidor de video bajo demanda y el cliente multimedia, tal como se explicó en 2.3.5.
- ✓ **Reproductor Multimedia:** es necesario para la reproducción de los contenidos multimedia difundidos bajo streaming.
- ✓ **Interfaz Gráfica de Usuario:** (GUI, *Graphical User Interface*): muestra de forma organizada la información respectiva sobre los contenidos multimedia disponibles.

Los servidores de seguridad y facturación no se consideran dentro del diseño de esta solución piloto, pero podrían ser integrados en una solución más robusta a futuro.

Para integrar al servicio de VoD, una interfaz por voz que haga uso de VoIP, se identificó que eran necesarios los siguientes elementos:

- ✓ **Motor de reconocimiento:** su función es recibir la voz, procesarla, identificar las palabras y entregarlas en un formato que puedan ser entendidas y utilizadas por los componentes del sistema.
- ✓ **Terminal de Usuario de VoIP:** debe recibir la voz cuando el usuario habla, codificarla y enviarla hacia el motor de reconocimiento. Dicho terminal puede estar a cargo de un softphone o un teléfono IP.
- ✓ **Servidor de Comunicaciones IP:** permite la comunicación entre el cliente multimedia y el motor de reconocimiento.

Una vez se han identificado y definido los elementos básicos necesarios para construir una solución de interactividad basada en VoIP para el manejo del servicio de VoD, el siguiente paso es definir cómo dichas funciones se desempeñarán y quién las realizará. De esta manera, las funciones del servidor de video, estarán a cargo de un servidor de video *Streaming*.

Dado que algunos STB en un ambiente en IPTV, acceden a los servicios interactivos por medio de páginas web, solicitadas desde el navegador integrado en el mismo, se considera que el servidor de aplicaciones estará a cargo de un servidor web. Así mismo, el acceso desde el terminal de usuario se realiza a través de un navegador web y la GUI de la solución se encuentra restringida a ser una página web.

La solución que se propone dentro del presente proyecto de grado, es un acercamiento inicial y piloto, por tanto, se ha elegido la manera más básica y sencilla de implementarla por medio de una arquitectura de organización de servidores centralizada.

Para dar cumplimiento con el requerimiento de modularidad, cada uno de los elementos identificados debe ser soportado por un módulo independiente dentro del esquema que se propone. Luego de un análisis de los requisitos y elementos expuestos anteriormente, en la Figura 8, se muestra el esquema general de la solución piloto propuesta.

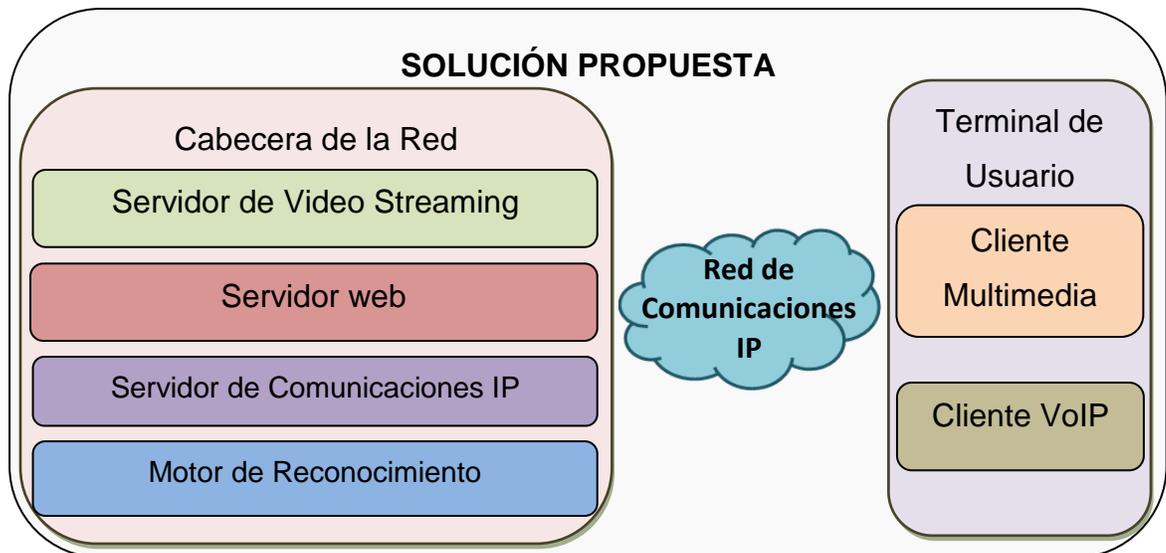


Figura 8. Esquema modular de una solución de interactividad.

El esquema general cuenta con 6 componentes, denominados: módulo de servidor de video streaming; de servidor web, de comunicaciones IP; motor de reconocimiento; cliente multimedia y de Voz sobre IP (VoIP, *Voice over IP*). La forma de comunicación entre los diferentes elementos de la solución, figura 8, se hará por medio de protocolos estandarizados, que se elegirán y explicarán en la siguiente sección.

Dentro de la solución, la red de comunicaciones permite la comunicación entre los elementos de la cabecera de red y del terminal de usuario. La red de comunicaciones debe soportar el transporte de información sobre el protocolo IP y puede ser implementada sobre cualquiera de los tipos de medios de transmisión explicados en el capítulo 2.

3.5 IMPLEMENTACIÓN, CONFIGURACIÓN Y PUESTA A PUNTO

Dentro de esta fase, ver figura 6 sección 3.2, se procede a configurar e implementar cada uno de los módulos pertenecientes al esquema propuesto, tal como se explicó anteriormente.

3.5.1 Actividad 1: Configuración de los Servidores de Video y Web

Es la primera actividad a desarrollar, ver figura 7, en la cual se elige, instala, configura y verifica el correcto funcionamiento de los módulos de servidores de video *streaming* y web.

3.5.1.1 Servidor de Video Streaming

La función del servidor de video streaming es enviar los flujos de video y permitir las funciones de control sobre los mismos (reproducir, pausar, detener, etc.).

Para elegir el servidor de video streaming a instalar, se analizan las opciones disponibles actualmente en el mercado, en la Tabla 14 y 15, se sintetizan los principales servidores de video streaming bajo licencia privativa y libre, respectivamente. Para estos se describen los sistemas operativos, códec y protocolos soportados.

Tabla 14. Servidores de Video Streaming bajo licencia privativa

Nombre	Descripción	Sistema operativo	Códec	Protocolos soportados
Unreal Media Server	Servidor de video <i>streaming</i> que soporta hasta 15 conexiones simultáneas.	Windows	MPEG layer 3, THEORA, VORBIS, WMV, WMA	HTTP
Wowza Media Server 2	Servidor multimedia flash que permite hasta 10 conexiones simultáneas, sólo si se adquiere la licencia para desarrolladores.	Windows, Linux, Mac OS X y Unix.	H.264	RTMP, RTSP, RTP, HTTP, NFS/AFS, entre otros.
Flash Media Streaming Server 3.5	Propiedad de Adobe Systems, utilizado para VoD, Video en vivo y comunicaciones en tiempo real.	Windows, Linux y Mac OS X.	H.264	RTMP, HTTP
Windows Media Services	Servidor de <i>streaming</i> multimedia, incorporado en el sistema operativo Windows Server.	Windows	WMV	HTTP, RTSP, MMS, RTP, UDP, TCP.
Quick Time Streaming Server	Es una herramienta asociada al sistema operativo Mac-OS que facilita la transmisión de videos a través de una interfaz de usuario sencilla.	Mac OS	H.264,H.263, MPEG-4 AVC, MPEG-4 PARTE 2	RTP/RTSP

Tabla 15. Servidores de Video Streaming bajo licencia libre.

Nombre	Descripción	Versión	Sistema operativo	Códec	Protocolos soportados
Peercast	Herramienta libre que hace uso de la transmisión peer to peer, para que el usuario sea un proveedor de contenidos.	0.1218 (17-12 - 2007)	Windows, Linux y Mac OS X	MPEG layer 3, THEORA, VORBIS, WMV, WMA	HTTP
Darwing Streaming Server	Servidor basado en el código de <i>Quicktime Streaming Server</i> , disponible bajo licencia de software libre.	6.03 (10 - 05 - 2007)	Windows, Linux, Mac OS X y Unix.	H.264,H.263, MPEG-4 AVC, MPEG-4 PARTE 2	RTP, RTSP, RTCP
Flumotion	Servidor de <i>streaming</i> , disponible para la difusión de contenidos en vivo y bajo demanda.	0.61 (09 - 09 - 2009)	Windows, Linux y Mac OS X	THEORA y VORBIS	HTTP
Hélix DNA server	Perteneciente a Real Networks, éste servidor envía contenidos multimedia bajo demanda a gran variedad de dispositivos.	11.1 (10 - 06 - 2006)	Windows, Linux y Unix.	REAL VIDEO, REAL AUDIO y MPEG LAYER 3	HTTP, RTSP, RTP, TCP, UDP, RDT
VLC	Es el producto estrella de la empresa videoLAN, con versiones para servidor de video streaming y reproductor multimedia.	1.0.3 (10 - 2009)	Windows, Linux, Mac OS X y Unix.	MPEG-1 VIDEO MPEG-2 VIDEO y AUDIO MPEG-4 VIDEO y AUDIO WMV 1/2 H/I 263 H.264, AVC, entre otros.	HTTP, RTSP, RTP, UDP, MMSH
RED5	Implementación java de Flash Media Server, basada en ingeniería inversa de los protocolos RTMP y AMF.	0.9.0 rc2 (03 - 11 - 2009)	Windows, Linux	H.264, MPEG LAYER 3	RTMP
OMSP - Open Media Streaming Project	Es un proyecto de software libre, para el streaming de contenidos multimedia sobre el protocolo IP.	-	Linux	MPEG 1/2, MPEG-4 PARTE 2, H.264, H.263, H263+, THEORA, VORBIS, AAC, MPEG LAYER 3.	RTP Y RTSP

Para la elección del servidor de video streaming adecuado, se definen los siguientes criterios:

- ✓ Soporte a los protocolos RTP, RTSP, RTCP, necesarios para el transporte de los contenidos en tiempo real. De esta manera, se cumple el requerimiento acerca del soporte a protocolos estandarizados.
- ✓ Soporte a diversidad de códec recientes, que posibiliten la oferta de contenidos en formatos de alta calidad con bajo consumo de recursos de red.
- ✓ Robustez.
- ✓ Capacidad de integración con otros servidores.
- ✓ Documentación disponible.

Se considera que Darwin Streaming Server (DSS), cumplen con los requerimientos expuestos. El proceso de instalación y configuración del mismo, se presenta en el Anexo A.

Un inconveniente de DSS, es el proceso adicional de incorporación de pistas de indicaciones al contenido multimedia. Éste ocurre previo a la adición de los videos al servidor y es necesario para la transmisión de los mismos en tiempo real.

Gracias a la interfaz web integrada a DSS, éste se configura y gestiona de manera sencilla, facilitando la adición de contenidos fácilmente. Además, ésta presenta información acerca del tiempo de conexión, tasa de transferencia y pérdida de paquetes de las sesiones establecidas

Para verificar el correcto funcionamiento del servidor de video streaming, se solicita un video desde varios reproductores multimedia como VLC, Quick Time Player, y Tótem, a quien el servidor responde de manera satisfactoria, enviando los contenidos multimedia solicitados y permitiendo las funciones de control sobre los mismos (pausar, reproducir y detener).

Los contenidos multimedia que difunde DSS, utilizan el códec H.264, debido a que éste permite entregar imágenes de alta calidad, precisando de pocos recursos de almacenamiento y transporte. Además, se encuentra estandarizado, asegurando el soporte del mismo por variedad de reproductores multimedia.

3.5.1.2 Servidor Web

El siguiente módulo dentro de la primera actividad, es el servidor web, quien aloja la guía de contenidos e intercomunica el cliente multimedia y el servidor de video streaming previo al establecimiento de la sesión necesaria para la transmisión del video.

La Tabla 16, sintetiza las principales características de algunos servidores web disponibles actualmente en el mercado tanto para licencias privativas como libres.

Para elegir el servidor web adecuado, los criterios a tener en cuenta son el desempeño, eficiencia, robustez, facilidad de integración con otros servidores y amplia documentación acerca del mismo. Acorde a lo anterior, se elige Apache para el módulo de servidor web.

Posterior a la instalación del servidor web, se verifica el correcto funcionamiento del mismo, accediendo desde un navegador a una de las páginas web previamente almacenadas dentro de este.

Tabla 16. Servidores web disponibles actualmente.

Servidor	Descripción	Versión	Sistema Operativo	Protocolos soportados	Licencia
Apache	Es el servidor web más utilizado a nivel mundial, y tiene como objetivo proveer servicios HTTP de forma segura y eficiente.	2.2.14	Todos	HTTP, PHP, TCL, PERL, PHYTON.	Libre
Apache TomCat	Es una implementación software de las tecnologías Java Servlet y JavaServer Pages para implementar un servidor web.	6.0.20	Todos	JSP, PHP.	Libre
Sun Java Web Server	Es uno de los servidores web líderes en el mundo, diseñado para medianas y grandes empresas. Entrega una infraestructura completa para aplicaciones web.	7.0	Solaris, Windows, Linux	JSP, PHP, HTTP	Privativa
Zeus	Este servidor permite crear, manejar y entregar contenidos web. Además, los clientes pueden visualizar y manipular el flujo de tráfico presente en sus aplicaciones web.	4.3	Todos	PHP, HTTP, JSP.	Privativa
Roxen	Es un servidor web de fácil configuración, con arquitectura modular, e integrado con MySQL.	5.0.403	Todos	HTTP, JSP, PHP.	Libre
AOL Server	Es un servidor web multihilo, con habilitador TCL y recomendado para sitios web dinámicos y a gran escala.	4.5.1	Todos	HTTP	Libre
Microsoft Windows Server	Sistema operativo dedicado completamente a trabajar como servidor, dentro de éste se encuentra integrado un servidor web.	v2008	Windows	FTP, HTTP	Privativa
Cherokee Web Server	Es muy rápido, flexible y fácil de configurar, compatible con diversas tecnologías como: balanceo de bases de datos, proxy HTTP, video <i>streaming</i> , etc.	0.99.39	Linux, Windows, Solaris	HTTP, PHP, JSP	Libre

La siguiente etapa de la primera actividad, es la integración del servidor web con el servidor de video streaming, proceso que se lleva a cabo por medio de los siguientes pasos:

1. Instalación del **complemento de cliente de VLC**⁴ en el navegador web: éste habilita la visualización de contenidos difundidos por video streaming a través de una página web.
2. Creación de la página web: para habilitar el uso del complemento de VLC, ésta se debe crear de acuerdo a las recomendaciones disponibles en el sitio web de su creador⁵. Dentro de ésta se debe incluir el Localizador de Recursos de Medios (MRL, *Media Resource Locator*), que tiene un formato similar a una URL e indica la ubicación del contenido multimedia en el servidor de video.
3. Solicitud de la página: a través del navegador se solicita la página web que ha sido creada, quien a su vez solicita el contenido multimedia de acuerdo a la dirección del MRL.
4. Reproducción: finalmente, dentro del navegador inicia la reproducción del contenido solicitado, comprobando la integración del servidor web y el servidor de video streaming.

NOTA: al inicio de la reproducción, se presentan pequeños retardos causados por la decodificación del video, sin embargo, luego de unos segundos la reproducción continúa normalmente.

Con la verificación de la integración de los servidores de video streaming y web, concluye la primera actividad. A continuación se explica la segunda actividad dentro de la fase de configuración en implementación.

3.5.2 Actividad 2: Instalación y Configuración del Servidor de Comunicaciones IP y el Motor de Reconocimiento

Durante la segunda actividad, ver figura 7 - sección 3.2, el servidor de comunicaciones IP y el motor de reconocimiento se instalan y configuran. Del mismo modo, la gramática que utiliza la solución y la interfaz entre el cliente y el servidor de comunicaciones se implementan. Esta actividad concluye con la verificación del correcto funcionamiento de los mismos.

3.5.2.1 Servidor de Comunicaciones IP

Entre las distintas opciones disponibles en el mercado [63], [64], [65], [66], [67], se encontró que las más robustas, estables, con diversidad de aplicaciones y que ofrecían mejores opciones de interconexión con otros programas, son las que se describen en la tabla 17; entrese selecciona Asterisk, por su familiaridad de uso, fácil instalación y

⁴Disponible:http://wiki.videolan.org/VLC_media_player

⁵Disponible:<http://wiki.videolan.org/Documentation:WebPlugin>

configuración, por la variedad de lenguajes de interconexión soportados y los módulos externos con los que se puede integrar, finalmente, y el más importante, por la variedad de motores de reconocimiento de voz con los cuales se puede conectar.

Tabla 17. Servidores de comunicaciones IP disponibles actualmente.

Característica	Servidor			
	Asterisk	FreeSwitch	CallWeaver	Yate
Última versión	Asterisk 1.6.2.14 Asterisk 1.8	1.06	1.2.1	3
Licencia	GPL	Mozilla Public License	GPL	GPL
Usos	VoIP gateways, call center ACDs y sistemas IVR.	Softwitch, PBX	PBX(basada en Asterisk 1.2)	Gateway VoIP-PSTN, servidor de telefonía IP
Codecs	G.722.1, G.722, G.711, G.726, AAL2, G.726, G.723.1, G.729A, ADPCM, iLBC, Speex, LPC-10, slin, GSM. Video Codecs: H.261, H.263, H.263+, H.264.	CELT, G.722.1, G.722, G.711, G.726, AMR, iLBC, Speex, entre otros.	AMR-NB, AMR-WB, G.722.1, G.722.2, G.723.1, G.728, G.729, G.729.1	GSM, Speex NB, WB y UWB, iLBC, AMR-NB
Protocolos	H.323, SIP, MGCP, ySCCP, IAX, SRTP, IPv4/IPv6, JABBER, XMPP, Jingle, multicast RTP, entre otros	SIP, UDP, TCP, SCTP and TLS, SRTP, IPv4/IPv6, ZRTP, Jingle H.323, SCCP, entre otros.	H.323, IAX2, MGCP y SIP y SCCP, RTP, TCP/TLS and SRTP	H.323, IAX2, MGCP y SIP, SS7/ISDN, Jingle, RTP.
Integración con motores de reconocimiento	Soporte nativo y con MRCP (lumenvox, Verbio, Loquendo, Sphinx,)	Soporte nativo y con MRCP (Lumenvox y Vestec, Nuance)	Mediante PocketSphinx	Mediante PocketSphinx + gstreamer
Desempeño	Se han realizado diversas pruebas de desempeño y se han obtenido muy buenos resultados. En [67] se muestran los resultados de una de las pruebas, en la cual se muestra como conclusión, que un servidor con	De acuerdo con el desarrollador del servidor de comunicaciones, el desempeño varía dependiendo de la aplicación.	Mucho más rápido y eficiente en la ejecución del DialPlan.	No existe información disponible al respecto.

Análisis del Uso de Interactividad basada en Voz para el manejo del Servicio de Video bajo Demanda

	Asterisk, por cada 1 GHz de capacidad de procesamiento puede manejar 100 llamadas simultáneas sin translación de Codecs (utilizando G.711)			
Sistema Operativo	GNU/Linux (x/86 y PPC), OpenBSD, FreeBSD, y Mac OS X.	Windows, Mac OS X, Linux, Solaris y *BSD	Linux, FreeBSD, NetBSD, OpenBSD, MacOS X, Open/Solaris	Linux, FreeBSD, Windows.
Lenguajes	AGIs(Java, C/C++, PHP, .NET, Perl, Python, ActiveX, Pascal, ruby)	JavaScript, Python, Perl, Lua, C /C++,java, .NET	Mediante OGIs, similar a las AGIs de asteriks	C++, PHP y Python.
Aplicaciones	Integración con web services, Correo de voz, conferencia, call parking, lector RSS, soporte de FAX sobre IP, call center, integración con sistemas CRM, presencia, entre otros,	Correo de voz, colas, conferencia, call parking, lector RSS, soporte de FAX sobre IP, entre otros,	IVR, conferencias, colas, Soporte de fax sobre IP	Conferencia, call parking, colas, IVR.

La función de Asterisk, es comunicar la aplicación cliente con el motor de reconocimiento, para que éste pueda identificar los comandos entregados a través de la voz. Dicha comunicación, se realiza a través de una llamada y se establece por medio de los protocolos de señalización.

La Tabla 18, sintetiza los principales protocolos de señalización soportados por Asterisk, entre estos, se elige Protocolo de Inicio de Sesión (SIP, *Session Initiation Protocol*), debido a su facilidad de configuración y el soporte del mismo por parte de diversos terminales de usuario.

Además de los protocolos de señalización, es necesario emplear códec de muestreo y compresión de la voz, para el transporte de la misma durante la llamada.

La Tabla 19, sintetiza algunas características de los principales códec soportados por Asterisk. Entre estos se elige el códec G.711, debido a que es menos sensible a la pérdida de paquetes, presenta menos retardo de decodificación, y entrega la voz al motor de reconocimiento con buena calidad en comparación con los otros códec, permitiendo el reconocimiento de los comandos más fácilmente [72], [73].

Tabla 18. Protocolos de señalización asociados a VoIP.

Protocolo	Características
H.323	Es un protocolo de transporte y señalización estandarizado por la Unión Internacional de Telecomunicaciones (ITU, <i>International Telecommunications Union</i>) para sistemas de comunicaciones multimedia basados en paquetes [68]. Inicialmente diseñado para el servicio de video conferencia, actualmente, es ampliamente empleado en comunicaciones de VoIP [69]. Tecnológicamente está mejor diseñado que SIP, sin embargo, dada su complejidad, está siendo reemplazado gradualmente. A diferencia de SIP, H.323 no tiene problemas con los firewalls, lo que le permite atravesar diferentes redes [69].
SIP	Es un protocolo de señalización propuesto por la IETF, para la transmisión de videos, música y contenido multimedia en tiempo real [69]. En comunicaciones de voz sobre IP, SIP se encarga del establecimiento, control y finalización de la llamada. El transporte de la voz se realiza a través de paquetes RTP. SIP tiene algunos problemas con la Traducción de Direcciones de Red (NAT, <i>Network Address Translation</i>) y cruzando firewalls [70].
IAX2	IAX2 (<i>Inter-Asterisk eXchange</i>) es un protocolo abierto diseñado para la comunicación entre centrales Asterisk, éste ha sido modificado y en este momento se encuentra a lo largo de la cadena de enlace de VoIP [69]. IAX2 combina los paquetes de señalización y de voz en un mismo flujo de datos, convirtiéndose en un protocolo de señalización y transporte [69], [70]. Contrario a los anteriores protocolos, IAX2 ofrece varias formas de encriptar la información y funciona perfectamente en redes con altos niveles de seguridad [69].
XMPP	XMPP (<i>eXtensible Messaging and Presence Protocol</i>), es otro protocolo desarrollado por la IETF para VoIP [70], [71]. Para establecer una llamada, es necesario un servidor para direccionar las llamadas y una pasarela que permita la comunicación con otras redes [71].
MGCP	El Protocolo de Control de Pasarela de Medios (MGCP, <i>Media Gateway Control Protocol</i>) es estandarizado por la IETF. Funciona bajo el modelo maestro-esclavo y precisa de un controlador externo para permitir la comunicación entre dos teléfonos [69]. Se considera un protocolo complejo, razón por la cual se recomienda utilizar SIP [69].

Tabla 19. Códec Asociados a VoIP soportados por Asterisk.

CÓDEC	Tasa de Bits (Kbps)	Comentarios
G.711	64	También se conoce como Modulación por Pulsos Codificados (PCM, <i>Pulse Code Modulation</i>). De éste se derivan otros códec y no requiere licencia para su uso [69].
G.722	48, 56 o 64	Utiliza Modulación por Pulsos Codificados Diferencial Adaptativa de sub-banda (SB-ADPCM, <i>Sub-band Adaptive Differential Pulse Code Modulation</i>). Para la transmisión de la señal, el audio se divide en dos sub-bandas, las cuales se codifican individualmente y se transmite a diferentes velocidades, dependiendo de la tasa de codificación [74].
G.723.1	5,3 o 6,3	Es un códec de muy bajo ancho de banda, utilizado en conjunto con H.323. Éste requiere una licencia en el decodificador [69].
G.726	16, 24 o 32	Se conoce como Modulación por Pulsos Codificados Diferencial Adaptativo (ADPCM, <i>Adaptive Differential Pulse-Code Modulation</i>). Ofrece la misma calidad que G.711 ocupando la mitad del ancho de banda. Una de las ventajas del G.726 es la baja carga de procesamiento que requiere [69].

G.729A	8	En contraste con el bajo ancho de banda utilizado, este códec entrega alta calidad. Como desventajas se encuentra el alto grado de procesamiento en el decodificador y la compra de una licencia necesaria para su uso [69].
GSM	13	Entrega excelente calidad, con poco ancho de banda y carga de procesamiento [69].
iLBC	13,3	El iLBC (<i>Internet Low Bitrate Codec</i>), cuenta con un equilibrio entre el ancho de banda y la calidad que ofrece, aunque requiere altos niveles de procesamiento [69].
Speex	Variable	Es un códec que varía la tasa de transmisión como respuesta a las características de la red y la calidad de la voz requerida [69].

Luego de elegir el protocolo de señalización y el códec de voz a emplear, se instala Asterisk como servidor de comunicaciones IP, y se modifican los archivos *sip.conf*, *extensions.conf* y *modules.conf* necesarios para el correcto funcionamiento de la solución (Anexo A).

Dentro del archivo *sip.conf*, se configuran los parámetros relacionados con la señalización de SIP, así como los usuarios que podrán acceder al servidor de comunicaciones.

extensions.conf, es el archivo más importante, debido a que dentro de este se establece la lógica de funcionamiento del servidor. Éste contiene la información acerca del comportamiento de una llamada, y es independiente para cada extensión disponible. Así mismo, en éste se configura un umbral cercano a 1000, que le indica al motor de reconocimiento cual debe ser el grado de exactitud en el reconocimiento de las palabras.

El servidor de comunicaciones y el motor de reconocimiento, se comunican de acuerdo a los parámetros establecidos en el archivo *modules.conf*.

Para verificar el funcionamiento del servidor de comunicaciones IP, desde un softphone⁶ se realiza una llamada a una de las extensiones, comprobando mediante un mensaje de eco, que éste responde correctamente.

La comunicación de Asterisk con sistemas externos, necesario para la integración con los otros módulos de la solución, se realiza por medio de la Interfaz a la Pasarela Asterisk (AGI, *Asterisk Gateway Interface*). Ésta se encuentra disponible en diversos lenguajes de programación, entre los que se encuentran JAVA, PHP, PERL, C, C++, entre otros.

Para la solución de interactividad propuesta, se elige la AGI Asterisk-JAVA⁷, debido a que ésta permite el acceso a las variables y funciones de control de llamadas de Asterisk de manera sencilla y rápida.

⁶Un softphone es una aplicación instalada en un computador, utilizada para realizar y/o recibir llamadas.

⁷Disponible: <http://asterisk-java.org/1.0.0.M3/>

3.5.2.2 Motor de Reconocimiento

Un motor de reconocimiento funciona bajo modelos estadísticos y requiere entrenamiento durante su etapa de diseño. A través del entrenamiento se construyen modelos acústicos y del lenguaje, que posteriormente serán utilizados para comparar las entradas de los usuarios.

El motor de reconocimiento como parte de la solución, debe recibir la voz, identificar las palabras pronunciadas y entregarlas en un formato que sea entendible por los otros módulos. La Tabla 20, presenta los principales motores de reconocimiento, integrables con Asterisk, independientes del locutor y sin entrenamiento por parte del usuario.

Tabla 20. Motores de Reconocimiento.

Motor de Reconocimiento	Características	Licencia
LumenVox ⁸	Está orientado a aplicaciones telefónicas o con arquitectura basada en teléfonos. Esta directa y perfectamente integrado con la plataforma Asterisk y al Dial Plan mediante un puente conector único de Digium (empresa creadora de Asterisk).	Privativa
CMU Sphinx ⁹	Es una colección de sistemas de reconocimiento del habla y desarrollados por la universidad Carnegie Mellon. Se destaca por su versatilidad y porque puede o no funcionar en tiempo real dependiendo de la versión empleada. La integración con Asterisk puede realizarse a través de un complemento, que funciona bajo el modelo cliente/servidor.	Licencia BSD
Verbio ¹⁰	Es un sistema de reconocimiento y síntesis de voz (ASR y TTS), que hace uso de gramáticas preestablecidas o creadas por los usuarios. Para la integración con Asterisk, se han desarrollado nuevas aplicaciones que se comunican directamente con ésta sin la intervención de una AGI.	De evaluación y Privativa
Loquendo ¹¹	Loquendo es el único proveedor de tecnologías vocales, que garantiza una amplia gama de voces e idiomas de alta calidad para cualquier entorno y sistema. Para la integración con Asterisk, se utiliza el API Asterisk Loquendo, que es un conjunto de funciones que simplifican el proceso de creación y gestión de diálogos.	Privativa

Entre los motores de reconocimiento de la tabla 20, se elige LumenVox, debido a que hace parte de la familia de productos de Asterisk, lo que asegura alta compatibilidad y reducción de la probabilidad de error durante la comunicación. Aunque este motor de

⁸ Disponible: <http://www.lumenvox.com/>

⁹ Disponible: <http://cmusphinx.sourceforge.net/>

¹⁰ Disponible: <http://www.verbio.com/webverbio3/html/index.php>

¹¹ Disponible: <http://www.loquendo.com/es/>

reconocimiento está disponible bajo licencia privativa, ésta es de bajo costo y fácil adquisición.

El motor de reconocimiento de LumenVox, hace uso de gramáticas, las cuales contienen las palabras que éste puede identificar y que el sistema está en condición de entender. Ésta se encuentra almacenada en un archivo y es consultado cada vez que se recibe una solicitud de reconocimiento.

En la solución se emplean 4 gramáticas, las cuales se activan y desactivan dependiendo del proceso que esté realizando el usuario. Éstas emplean palabras de uso común, las cuales han sido establecidas por medio de métodos informales de evaluación a través de entrevistas cortas, con preguntas acerca de cuáles son las palabras más utilizadas durante la interacción con el DVD (*Digital Video Disc*). Las gramáticas empleadas en la solución propuesta, se encuentran en el Anexo A.

El proceso de instalación y configuración del motor de reconocimiento, se describe en el Anexo A. Para verificar funcionamiento del mismo, se establece una llamada a una de las extensiones de Asterisk, quien a su vez se conecta con el motor de reconocimiento y le entrega la voz, quien la reconoce y la muestra en consola.

En conjunto con la palabra, el motor de reconocimiento muestra el grado de exactitud con la que ésta fue reconocida.

Una vez se ha verificado la integración del servidor de comunicaciones con el motor de reconocimiento se da por terminada de forma exitosa la segunda actividad.

3.5.3 Actividad 3: Despliegue del Cliente e Integración del Sistema

En esta actividad, ver figura 7 sección 3.2, se configura el cliente VoIP y el cliente multimedia y se realiza la integración con los otros módulos de la solución propuesta.

3.5.3.1 Cliente VoIP

El cliente VoIP está a cargo de un teléfono IP, quien obtiene la voz, la codifica, la comprime y la envía hasta el servidor de comunicaciones IP. Dentro de la solución, se decide utilizar un softphone por su simplicidad de configuración e instalación. Además, a través de este se desea comprobar si el usuario desea o no, contar con un dispositivo palpable en el momento de entregar un comando.

Luego de la instalación del softphone, se le asigna una de las extensiones habilitadas dentro del servidor de comunicaciones IP, quien indica que se encuentra configurado correctamente y es posible establecer una llamada, a través de un mensaje de “usuario registrado” en la pantalla del teléfono.

3.5.3.2 Cliente Multimedia

El cliente multimedia es la aplicación que permite elegir, decodificar y reproducir el contenido multimedia transmitido por medio de streaming desde el servidor de video. Ésta se encuentra embebida dentro de una página web y puede ser solicitada desde el navegador web en un STB o computador.

La aplicación multimedia se divide en la GUI, y la interfaz entre el módulo de reconocimiento del habla y la GUI.

La Figura 9, muestra la GUI del cliente multimedia, quien presenta de manera organizada la guía de contenidos disponibles y adiciona información acerca de los mismos en la parte baja de la pantalla. Ésta ha sido diseñada tomando en cuenta las interfaces de varios proveedores de servicio de VoD.

Para el diseño de la GUI, también se consideran algunas directrices de diseño como tipo de letra, color de fuente, color de fondo, entre otros [75] - [78]. Así mismo, para dar cumplimiento a uno de los requerimientos de la solución, la interacción con la GUI se realiza por medio de comandos sencillos y cortos, disponibles visualmente en pantalla.

Con el fin de brindar realimentación al usuario, dentro de la GUI se adiciona un campo que indica el comando reconocido por el motor de reconocimiento.



Figura 9. Interfaz Gráfica de Usuario.

Posterior al diseño e implementación de la GUI, es pertinente su comunicación con el módulo de reconocimiento del habla y el servidor de video streaming, integrando el sistema en su totalidad.

La comunicación entre el módulo de reconocimiento del habla y la GUI, es posible realizarla a través de web sockets o Applets. Debido a su reciente publicación, los web sockets no se encuentran soportados por diversos navegadores web, lo que restringe el acceso a un determinado grupo de usuarios. De esta manera se elige trabajar con Applets para la integración la solución.

El Applet de la solución se encuentra escrito en JAVA y está embebido en la GUI y provee intercambio de información de manera fiable, ordenada y en tiempo real. Éste, sostiene constantemente una conexión con la AGI del módulo de reconocimiento del habla, a la espera de un nuevo comando.

Una vez un comando llega al Applet, éste solicita una nueva página web si el televidente ha cambiado de categoría. Si el comando es de control del flujo de video, el Applet lo ejecuta sobre el reproductor de video, quien a su vez envía el comando al servidor de video streaming.

Para verificar el funcionamiento de la solución, se solicita la GUI al servidor web, se establece una llamada con el servidor de comunicaciones IP y se procede a entregar comandos a través del softphone. Finalmente, por medio de cambios en la GUI y la entrega del video, se verifica la ejecución de los mismos.

La solución a la cual se llega luego de las actividades de la fase de configuración, implementación y puesta a punto, se muestra en la Figura 10.

El proceso interactivo inicia cuando el Cliente VoIP se registra en el servidor de comunicaciones IP y establece una llamada de manera permanente. El proceso de señalización de la llamada se realiza por medio del protocolo SIP y el transporte de la voz sobre RTP. Paralelo a dicho proceso, desde el STB se solicita la guía de contenidos al servidor web, para esto se hace uso del protocolo HTTP.

Una vez la llamada ha sido establecida y la guía de contenidos se muestra en pantalla, el usuario pronuncia un comando, que el cliente VoIP obtiene y envía al servidor de comunicaciones IP, éste lo pasa al motor de reconocimiento, quien lo reconoce y lo pasa a la API de java, para que encapsulado dentro de paquetes TCP viaje hasta el Applet del cliente y se ejecute.

Si el comando es un comando de navegación entre los contenidos de una categoría, el Applet lo ejecuta internamente. Si el comando es la solicitud de una nueva categoría, el STB solicita la página con la información necesaria al servidor web, comunicación que se realiza por medio del protocolo HTTP. Finalmente, si el usuario ha decidido ver una

película o realizar las funciones de “Trick Mode”, el STB envía la petición al servidor de video *Streaming*, haciendo uso de los protocolos RTP, RTCP y RTSP.

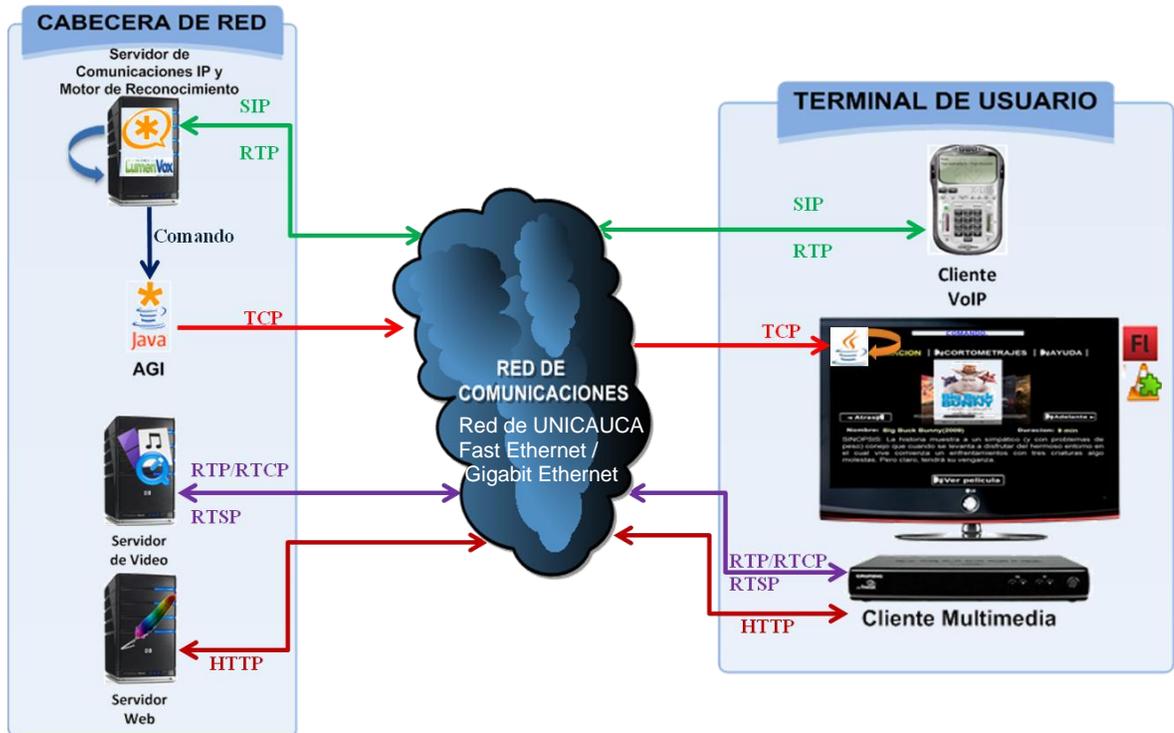


Figura 10. Solución de Interactividad.

El proceso de interacción con más detalle, se muestra en la Figura 11, y se explica a continuación.

1. El primer paso es que el cliente SIP realice el proceso de autenticación y establecimiento de sesión con el servidor de comunicaciones IP a cargo de Asterisk, quien permite el registro según lo establecido en el archivo de configuración “*sip.conf*”.
2. a. Mientras se establece la sesión SIP, el navegador web que reside en el STB, solicita al servidor web, la página con la guía de contenidos disponibles. La dirección IP a la cual se debe solicitar la página se encuentra pre-asignada en el navegador. El servidor web responde a dicha petición y
b. envía el archivo “*index.html*”, el cual posee dos marcos¹²: en el primero reside el Applet “*Applet.html*” y en el segundo se encuentra la introducción del servicio “*intro.html*”.
- c. Una vez la página se carga en el navegador, el Applet se ejecuta y permanece a la espera de comandos, los cuales sirven para controlar el segundo marco.

¹²Marco: permiten presentar documentos con vistas múltiples, que pueden ser ventanas o sub-ventanas independientes.

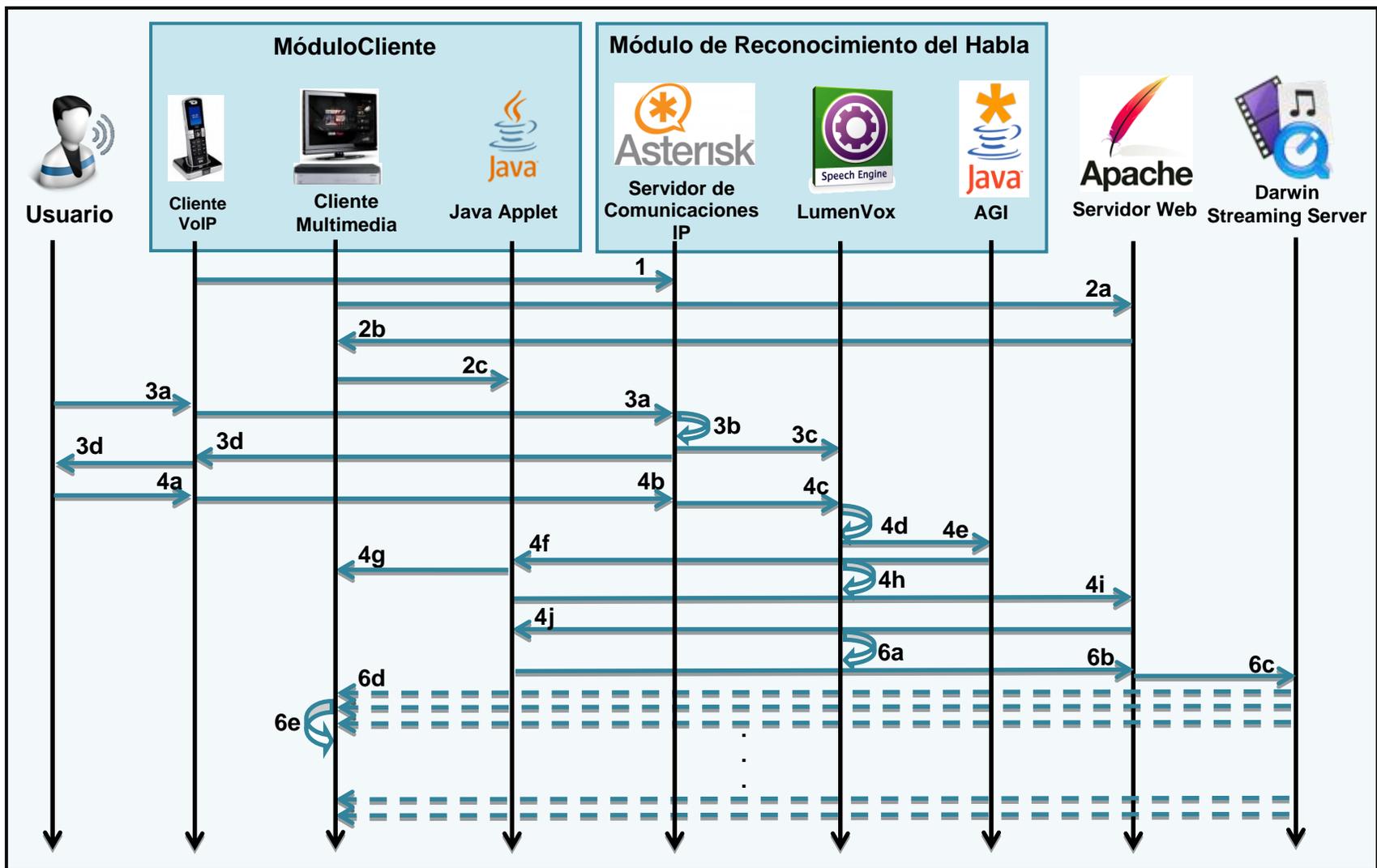


Figura 11. Proceso de Interacción.

3.
 - a. Cuando el usuario llama a una extensión telefónica preestablecida en el archivo de configuración “*extensions.conf*” del servidor Asterisk, ésta
 - b. contesta automáticamente,
 - c. crea la conexión con el motor de reconocimiento (Lumenvox), carga las gramáticas necesarias “*activar.grxml*”, “*inicial.grxml*”, “*menú.grxml*” y “*reproducción.grxml*” y activa “*activar.grxml*”. Posteriormente
 - d. entrega una señal para que el usuario pronuncie el comando.

4. Para iniciar el servicio el usuario debe decir “iniciar servicio”, el cual sigue el siguiente proceso:
 - a. El usuario pronuncia el comando.
 - b. El cliente multimedia lo obtiene y lo envía al servidor de comunicaciones IP.
 - c. Asterisk lo recibe y lo entrega al motor de reconocimiento.
 - d. El motor de reconocimiento lo identifica,
 - e. Lo envía a la AGI, ésta
 - f. Lo transmite al Applet java, quien
 - g. Lo muestra en pantalla y lo ejecuta, haciendo uso de las funciones JavaScript, almacenadas en “*voicevod.js*”, entonces,

5.
 - a. se desactiva la gramática “*inicial.grxml*” y se activa “*menú.grxml*”.
 - b. el Applet cambia el segundo marco, desactivando “*intro.html*” y
 - c. recibiendo “*base1.html*” desde el servidor web.

NOTA: Cualquiera que sea el comando que el usuario pronuncie, éste seguirá el proceso descrito en el numeral 4. La lista completa de los comandos se puede encontrar en las gramáticas del sistema en el Anexo A.

El archivo “*base1.html*” contiene la interfaz gráfica de una de las categorías de videos disponibles, y el usuario puede moverse entre los contenidos disponibles en esa categoría, ver la descripción de cada contenido e iniciar la reproducción, además se muestra el nombre de la categoría actual y de las otras categorías. Estando en este marco el usuario podrá decir comandos como: “Adelante” o “atrás” (para moverse de un contenido a otro); “Ver película”, “empezar película” o “iniciar película” (para iniciar la reproducción del flujo multimedia); “Animación” o “cortometrajes” (para cambiar de una categoría a otra). “*ayuda*” para que el sistema muestre en pantalla cuales son los comandos que puede pronunciar.

6. Para ver una película, el usuario pronuncia “*ver película*” o “*empezar película*” o “*iniciar película*”, el comando sigue el proceso detallado en el numeral 4, posteriormente
 - a. La gramática “*menú.grxml*” se desactiva y se activa “*reproducción.grxml*”.
 - b. Applet llama el archivo “*play.html*”, que se encuentra en el servidor web, y le entrega la dirección RTSP del contenido deseado así como el nombre del contenido y el nombre del archivo de categorías (“*base1.html*” o “*base2.html*”).

- c. Este se comunica con el servidor de video Streaming, para iniciar la transmisión del flujo de video,
- d. el STB lo recibe
- e. lo guarda en el buffer, lo decodifica y lo muestra en pantalla.

Durante la reproducción del video el televidente podrá entregar comandos como: *“Pausar”, “Iniciar”, “Aumentar volumen”, “Disminuir volumen”, “Silenciar”, “Categorías”* o *“menú películas”* (vuelve al menú de categorías para elegir un nuevo contenido, desactiva *“reproducción.grxml”* y activa *“menú.grxml”*).

En cualquier momento el usuario podrá detener la interfaz por voz, pronunciando *“desactivar reconocimiento de voz”*. Para activar nuevamente la interfaz el usuario deberá pronunciar *“activar reconocimiento de voz”*.

El funcionamiento interno del dial plan del servidor de comunicaciones y el proceso de activación y desactivación de las gramáticas del motor de reconocimiento, se describen en el anexo A.

Con la integración de la solución en su totalidad, se finaliza la fase de configuración, implementación y puesta a punto de la propuesta de interactividad del presente proyecto de investigación.

Dado que la solución que se propone dentro del presente proyecto de grado es una propuesta inicial, es necesario establecer cuál es el contexto de aplicación y cuáles serían los elementos necesarios para su puesta en funcionamiento.

3.6 CONTEXTO DE USO DE LA SOLUCIÓN

Dentro de esta sección se describen algunos ejemplos de escenarios de aplicación y los elementos necesarios para la puesta en funcionamiento de la solución piloto propuesta.

3.6.1 Contexto de Uso de la Solución

La propuesta de interactividad que se propone dentro del presente proyecto de grado, se encuentra orientada hacia ambientes de entretenimiento de bajo ruido, con pocos usuarios en el entorno de interacción (Ambiente residenciales y Habitaciones de Hotel). Los televidentes que se consideran harán uso de la solución, son personas que tengan capacidad de hablar claro, desde niños a adultos mayores, que se encuentran a una distancia promedio del televisor de entre dos y tres metros.

Un ejemplo de entorno de aplicación de la solución, es un sistema de video bajo demanda en un hotel, en el cual generalmente existen uno o dos usuarios en una habitación, lo que disminuye la probabilidad de que se entreguen comandos por error cuando se interactúa con otros televidentes. También, se considera que los hoteles son un posible campo de

aplicación, debido a que las habitaciones son un entorno aislado con baja interferencia por ruido externo.

Si se desea aplicar la propuesta del presente proyecto en el hogar, se recomienda integrar micrófonos con supresión de ruido y el usuario debe contar con una conexión a Internet de banda ancha, donde la velocidad mínima depende de la calidad de la imagen del video que se esté observando. Además debe reservar con su proveedor de red/servicio un mínimo de velocidad para la conexión entre el cliente VoIP y el servidor de comunicaciones IP.

3.6.2 Condiciones para la Aplicación de la Solución en un Entorno Real

La forma como tradicionalmente se ofrece el servicio bajo demanda y los elementos necesarios se observan en la figura 12, que corresponde con la infraestructura de red de Imagenio para la prestación de IPTV y de Video bajo demanda bajo ADSL [79].

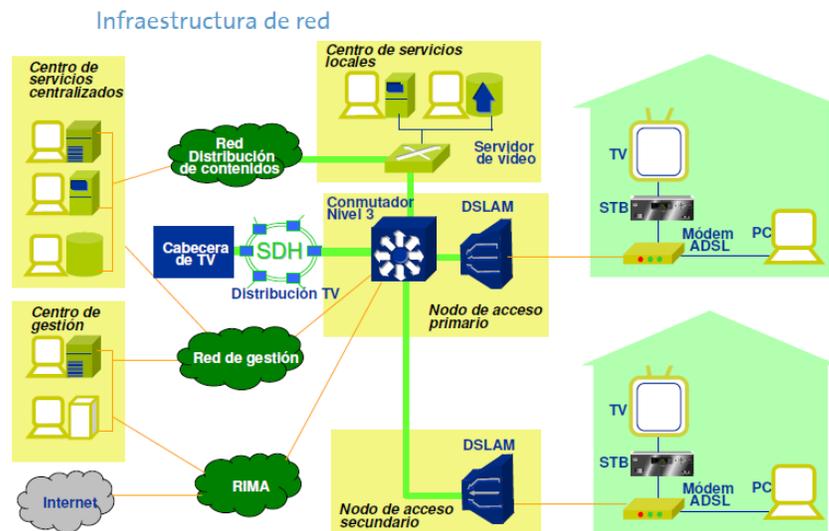


Figura 12. Infraestructura de red de Imagenio.

Para aplicar la solución piloto propuesta en el presente trabajo a dicha infraestructura se deberían agregar los siguientes elementos con las siguientes características:

1. En el domicilio del usuario:

- **Set Top Box:** dicho elemento deberá soportar: HTML, HTTP, HTTPS, DHTML, SSL, Cookie, Java® Script, Adobe® (Macromedia) Flash, formatos GIF, JPEG, PNG, Espial Escape®, Opera®, ANT®, Mozilla Browser, entre otros. RTP, RTSP, Para el envío de los contenidos. En cuanto al soporte de Codecs deberá soportar la decodificación de contenidos en H.264, MPEG-2, MPEG1-L3* (MP3), AAC, HE-AAC, Dolby® Digital 2.0/5.1 (AC-3), entre otros.

Por ejemplo: **IPTV STB-7710** y **STB-7711** de Siemens, **ADB-5721WNX** de Advanced Digital Broadcast, **CMS 1080 STMC**.

La **guía de contenidos** deberá estar basada en HTML, JavaScript, CSS, XML, etc.

- **Teléfono IP:** Podría ser remplazado por un teléfono IP integrado al control remoto, o al STB, dependiendo de fabricante del mismo. Dicho teléfono deberá establecer una sesión SIP con el servidor Asterisk ubicado en la cabecera de red.

2. Cabecera de red:

- **Servidor de comunicaciones Asterisk:** para este servidor se debe tener en cuenta que se establece una "llamada" para cada usuario que este interactuando con el servicio. Teniendo en cuenta lo anterior y dado que en un futuro podría existir una gran cantidad de usuarios accediendo simultáneamente al servicio por medio de la solución, se hace necesario el uso de un clúster de servidores, cuyo número de servidores variará acorde al dimensionamiento del sistema.

Para el dimensionamiento del número de servidores necesarios se pueden considerar los resultados obtenidos por [67], pruebas en las cuales se utilizó un servidor con las siguientes características:

Procesador Intel Xeon X3220 Quadcore, 2.40 GHz, 1066 MHz FSB, 2x4M cache, 4 GB RAM. El sistema operativo usado fue Red Hat Enterprise Linux Server release 5.1, x86-64. La tarjeta de red es una onboard GigE NIC.

Asterisk: 1.4.21.2.

Protocolo: SIP

Codec: G711.

Como resultado de dicho estudio, se estableció que un servidor alojando Asterisk, puede manejar 100 llamadas simultaneas (sin traducción de códec, usando G711 ulaw) por cada GHz de procesamiento.

En el momento de aplicar la solución piloto propuesta en el presente proyecto, se deberán realizar pruebas de carga del (los) servidor(es) para analizar la cantidad simultanea de llamadas que soporta servidor seleccionado.

- **Servidor de reconocimiento:** elemento esencial para la solución piloto, debido a que es quien realiza el proceso de reconocimiento. Para éste es esencial la cantidad de usuarios simultáneos que puede manejar el servidor en un determinado instante, lo que depende de:
 - ✓ El tamaño y complejidad de la gramática.
 - ✓ La longitud del audio.
 - ✓ La frecuencia de uso del sistema de reconocimiento.

La figura 13, muestra la relación entre el tiempo promedio de reconocimiento de voz y el número total de usuarios que pueden ser atendidos en un solo equipo, dependiendo del hardware del mismo [80].

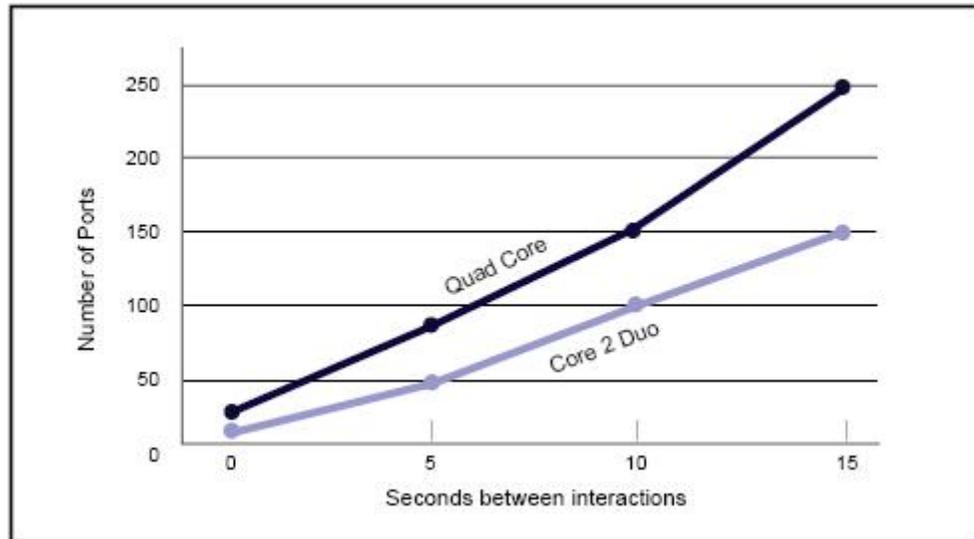


Figura 13. Comportamiento del motor de reconocimiento de LumenVox.

Teniendo en cuenta la figura 13, en el dimensionamiento de la solución en un entorno real, se debe evaluar cuál es el tiempo promedio de interacción entre un usuario y otros, de manera que sea posible determinar la cantidad de usuarios que pueden ser atendidos por un solo servidor y por ende la cantidad de servidores necesarios para la prestación del servicio.

- La **red de comunicaciones** tiene como único requerimiento el soporte al protocolo IP, dependiendo del proveedor del servicio elegir la forma bajo la cual desea realizar la conexión entre los servidores y el terminal de usuario. Algunas opciones disponibles para el despliegue de la red, se describen de manera general en el capítulo 2 [61].

Con la explicación de las condiciones necesarias para la realización de la solución en un entorno real, concluye el capítulo de descripción de la solución.

La solución que se propone, se genera a partir de un proceso incremental, a través de las fases de diseño, configuración, implementación y puesta a punto de la misma ya explicados. Por tanto, ahora es conveniente determinar que tan adecuada es una solución piloto de interactividad basada en voz para un servicio de VoD, proceso que se describe en el siguiente capítulo.

CAPÍTULO 4. EVALUACIÓN DE LA SOLUCIÓN PILOTO Y ANÁLISIS DE RESULTADOS

Dentro de este capítulo, se describen las evaluaciones realizadas a la solución de interactividad basada en voz para el manejo del servicio de VoD, descrita en el capítulo 3, con el objetivo de analizar si la voz es un mecanismo de interacción adecuado para el control de dicho servicio.

El primer enfoque de la evaluación considera la percepción del usuario frente a la propuesta (Evaluación Subjetiva), debido a que es el usuario quien hará uso de la interfaz en un ambiente real. Los resultados obtenidos de la evaluación de la solución piloto de interactividad basada en voz, serán comparados con los resultados obtenidos de la evaluación de un servicio de VoD controlado por control remoto. Además, ésta es valiosa en relación con la realimentación que se obtiene, la cual permite detectar las debilidades, fortalezas, y posibles factores susceptibles a reformas.

El segundo enfoque, describe el comportamiento de la solución desde la perspectiva de algunos parámetros de red (Evaluación Objetiva) como el retardo, el jitter y la pérdida de paquetes, permitiendo detectar posibles problemas en el procesamiento y transporte de la información, que puedan afectar la percepción del usuario.

4.1 ASPECTOS A EVALUAR

En esta sección se determinan los parámetros desde los cuales se evalúan la solución.

4.1.1 Enfoque Subjetivo

Para evaluar la solución desde un enfoque subjetivo, se construye el Cuestionario de Percepción del Servicio (CPS), que tiene como objetivo obtener datos cuantificables, después de la interacción del televidente con la solución propuesta.

El cuestionario incluye un total de 21 expresiones, positivas y negativas, con una escala tipo Likert¹³ de 5 niveles, en la cual, el usuario elige entre el nivel de conformidad o inconformidad con las mismas.

¹³ La escala tipo Likert permite obtener respuestas de forma rápida, mediante la selección de una respuesta con diferentes escalas de opinión.

Dado que no existe un cuestionario estándar para la evaluación de un servicio de VoD, se decide realizar la construcción de las expresiones, con base en los cuestionarios MUMMS (*Measuring the Usability of Multi-Media Software*) [81] y SASSI (*Subjective Assessment of Speech System Interfaces*) [82], ya que, estos permiten obtener la opinión de los usuarios, una vez se han realizados interacciones con sistemas multimedia y soluciones controladas por voz, respectivamente.

Con el fin de obtener la opinión de los usuarios sin causar molestia, por el número de preguntas a contestar, se decide integrar algunos de los enunciados presentes en los cuestionarios MUMMS y SASSI, considerando solo los de mayor relevancia en relación con la solución planteada en el presente proyecto de investigación

EL CPS, desea obtener la apreciación del usuario frente a la calidad de la solución propuesta, por esta razón, se considera la medida de los siguientes factores [81]:

- ✓ **Atractivo del sistema:** determina el grado de satisfacción y agrado de los usuarios frente al uso de la solución.
- ✓ **Grado de control:** obtiene la percepción del usuario, acerca de si la solución es fácil de controlar y responde en la forma en la que el televidente espera.
- ✓ **Eficiencia:** se refiere a si el sistema permite cumplir los propósitos de interacción que desea realizar el usuario, desde el punto de vista de la precisión y velocidad de respuesta de la solución, así como la carga cognitiva en el televidente.
- ✓ **Utilidad:** toma en cuenta la opinión del usuario, en concordancia a si el sistema le es útil y lo guía durante la interacción.
- ✓ **Facilidad de aprendizaje:** manifiesta la opinión del televidente, en relación con la facilidad de aprendizaje y uso de la solución.

El objetivo del presente proyecto es determinar si una interfaz por voz es apropiada para el manejo de un servicio de VoD. No obstante, se desea comparar la percepción del usuario cuando hace uso del control remoto y la sensación que genera un cambio en la entrega de comandos a través de la voz. De esta manera, el cuestionario de percepción del servicio, se aplica a dos interacciones, una a través del control remoto, emulando la entrega del servicio a través de un STB, y otra por medio de la solución propuesta, con la entrega de comandos a través de la voz.

El cuestionario de evaluación del sistema por voz, adiciona cuatro preguntas abiertas, con el fin de determinar la opinión del usuario frente a la solución y la elección entre la entrega de comandos a través del control remoto o la voz. El anexo B, contiene el cuestionario con las preguntas realizadas a los usuarios y las respuestas de los mismos. El análisis de los resultados obtenidos se muestran en la sección 4.3.3.

4.1.2 Enfoque Objetivo

Algunos parámetros a analizar desde el aspecto objetivo, son el retardo, jitter y pérdida de paquetes, los cuales permiten conocer parte del comportamiento de la solución desde la perspectiva de red.

Para medir los parámetros antes mencionados, es necesario el uso de las siguientes herramientas:

- ✓ **WireShark**¹⁴: programa utilizado para la captura y análisis de paquetes.
- ✓ **OmniPeek**¹⁵: aplicación empleada para obtener los valores de retardo, jitter y pérdida de paquetes del contenido multimedia, a partir de la captura realizada por WireShark.
- ✓ **Cace Pilot**¹⁶: aplicación empleada para obtener los valores de retardo, jitter y pérdida de paquetes de VoIP, a partir de la captura realizada por WireShark.

Una vez definidos los aspectos y herramientas para realizar la evaluación, a continuación se explica los escenarios en los cuales se lleva a cabo dicho proceso.

4.2 ESCENARIO DE EVALUACIÓN

En esta sección se explican el escenario de evaluación de la solución desde el aspecto subjetivo y objetivo.

4.2.1 Entorno de Evaluación

Dado que el entorno de evaluación no es real y será controlado, durante la evaluación se tratará de que éste sea lo más confortable y tranquilo posible, de manera tal, que no afecte el desarrollo de la prueba. Por tanto, se establece la oficina 448, de la Facultad de Ingeniería Electrónica y Telecomunicaciones (FIET), como el espacio/ambiente donde se realizará el proceso de evaluación de la solución de interactividad basada en voz, por ser un espacio con baja afluencia de personal.

La figura 14, muestra los elementos con los cuales, se realiza la evaluación de acuerdo al aspecto objetivo y subjetivo. En este caso, la red de comunicaciones se emula a través de la LAN de la FIET. Los elementos utilizados son:

- ✓ Televisor de 29”.
- ✓ 1 softphone (micrófono).

¹⁴ Disponible: <http://www.wireshark.org/>

¹⁵ Disponible: <http://www.wildpackets.com/products>

¹⁶ Disponible: http://www.cacotech.com/products/cace_pilot.html

- ✓ 1 computador para emular las funciones de un STB con conexión Ethernet: con las siguientes características: **disco duro** de 160 GB, **memoria RAM** de 3GB, **procesador** AMD turión 64x2 Mobile Techonology TL-58 de 1,59 GHz, **sistema operativo** Windows XP.
- ✓ Servidor con el servidor web, el servidor de video *streaming*, el servidor de comunicaciones IP y el motor de reconocimiento.
- ✓ Cable de interconexión entre el TV y el computador que emula el STB. Conversor de super video a RCA.
- ✓ Red de Comunicaciones: ésta será emulada a través de la Red de Área Local (LAN, *Local Area Network*) de la Universidad del Cauca, la cual esta soportada sobre tecnología Ethernet (Fast y Giga) e IP.

El servidor alberga los elementos de la cabecera de red de la propuesta de interactividad, descrita en el capítulo anterior. El computador portátil emula el STB y es el cliente multimedia, quien se conecta con el servidor web para acceder a la guía de contenidos, y recibe los contenidos multimedia desde el servidor de video streaming. Para enviar los comandos hacia el servidor de comunicaciones IP, se utiliza un softphone, quien hace las veces de cliente VoIP y se encuentra configurado en el computador portátil 2.

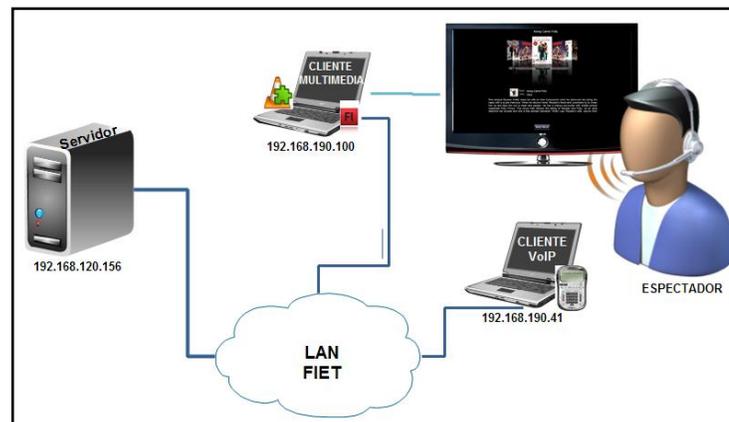


Figura 14. Escenario de Evaluación de la solución.

La figura 15, muestra los elementos utilizados para emular el manejo de un STB por medio del control remoto, entre los cuales se encuentra:

- ✓ Televisor de 29”.
- ✓ DVD Player con control remoto.
- ✓ DVD Video con la emulación del sistema de VoD.



Figura 15. Escenario de Evaluación con el Control Remoto

4.2.2 Descripción de la Población

Los usuarios que evalúan la solución se eligen de manera aleatoria, considerando diferente ocupación, edad, género y nivel educativo, debido a que un servicio de VoD, no se encuentra ligado a una población en particular. De acuerdo a Nielsen [83], [84], para obtener los problemas presentes en un sistema y la percepción del usuario frente a la solución son necesarios mínimo 5 participantes, de esta manera el número de los usuarios que evaluará la solución será superior a 5.

El tamaño de la muestra es de 21 personas, entre las que se encuentran estudiantes universitarios, estudiantes de educación media, profesionales en diferentes áreas, profesores, operarios de aseo, entre otros. La edad de los usuarios oscila entre 9 y 52 años, con un total de 7 mujeres y 14 hombres. En el anexo B, se describen la edad, género y ocupación de los usuarios que realizaron la evaluación de la solución propuesta.

4.3 EVALUACIÓN SUBJETIVA

En esta sección se describe el proceso de evaluación que se lleva a cabo y se analizan los resultados obtenidos a partir de la aplicación del CPS.

4.3.1 Proceso de Evaluación

Una vez se han establecido las herramientas y escenario para la evaluación de la solución, la Figura 16, muestra el proceso que se sigue para obtener la percepción de los usuarios luego de la interacción con el control remoto y la voz. Ésta consta de 4 etapas denominadas: introducción, interacción con el sistema, cuestionario post-prueba y análisis de resultados.



Figura 16. Proceso de evaluación.

4.3.2 Aplicación de la Evaluación

4.3.2.1 Introducción

Aplicando el proceso de evaluación descrito en la figura 14, inicialmente se explica al usuario lo concerniente al servicio de VoD. Luego, se describe el proceso de evaluación, el cual consiste de los siguientes pasos:

1. Elegir y controlar un contenido multimedia por medio del control remoto.
2. Entregar la opinión de la percepción del servicio por medio del CPS.
3. Elegir y controlar un contenido multimedia a través de la voz.
4. Opinar acerca de la solución por medio del CPS.

4.3.2.2 Interacción con el Sistema con control remoto

Una vez el usuario conoce el procedimiento a seguir, éste interactúa con el servicio. A partir de la observación se encuentra que el tiempo promedio de uso del servicio está comprendido entre 8 y 10 minutos. La interfaz gráfica utilizada para el servicio con control remoto se muestra en la figura 17. Ésta se realiza de manera similar a la GUI de la solución propuesta en el capítulo 3, con el objetivo de presentar al televidente el mismo servicio controlado por diferentes mecanismos de interacción.



Figura 17. Interfaz de Servicio de VoD a través del Control Remoto.

Luego que el televidente ha utilizado el servicio, éste responde durante unos minutos el CPS para el control remoto. El anexo B, contiene las respuestas de los usuarios referentes a esta interacción.

4.3.2.3 Interacción con la Solución Propuesta

Luego de responder el CPS para el control remoto, el usuario interactúa con la solución propuesta. Con base en la observación se percibe que el tiempo promedio de uso del servicio está comprendido entre 13 y 15 minutos. Posteriormente, el televidente responde las preguntas del CPS controlado por voz. El anexo B, contiene los cuestionarios y las respuestas de los usuarios.

4.3.3 Análisis de Resultados

En esta sección se realiza el análisis de los resultados obtenidos durante el proceso de evaluación descrito anteriormente, el cual busca obtener la percepción del usuario frente al uso de una solución de interactividad basada en voz para el manejo del servicio de VoD. Así mismo, se realiza una comparación con los resultados derivados de la interacción a través del control remoto.

En concordancia con los factores establecidos previamente, a continuación, se realiza un análisis del grado de atractivo, control, eficiencia, utilidad y facilidad de aprendizaje de la solución.

4.3.3.1 Atractivo del Sistema

La figura 18, presenta los resultados obtenidos de la aplicación del CPS referentes al aspecto atractivo del sistema controlado por medio del control remoto y la figura 19,

muestra los resultados de la aplicación del CPS para la solución de interactividad a través de VoIP.

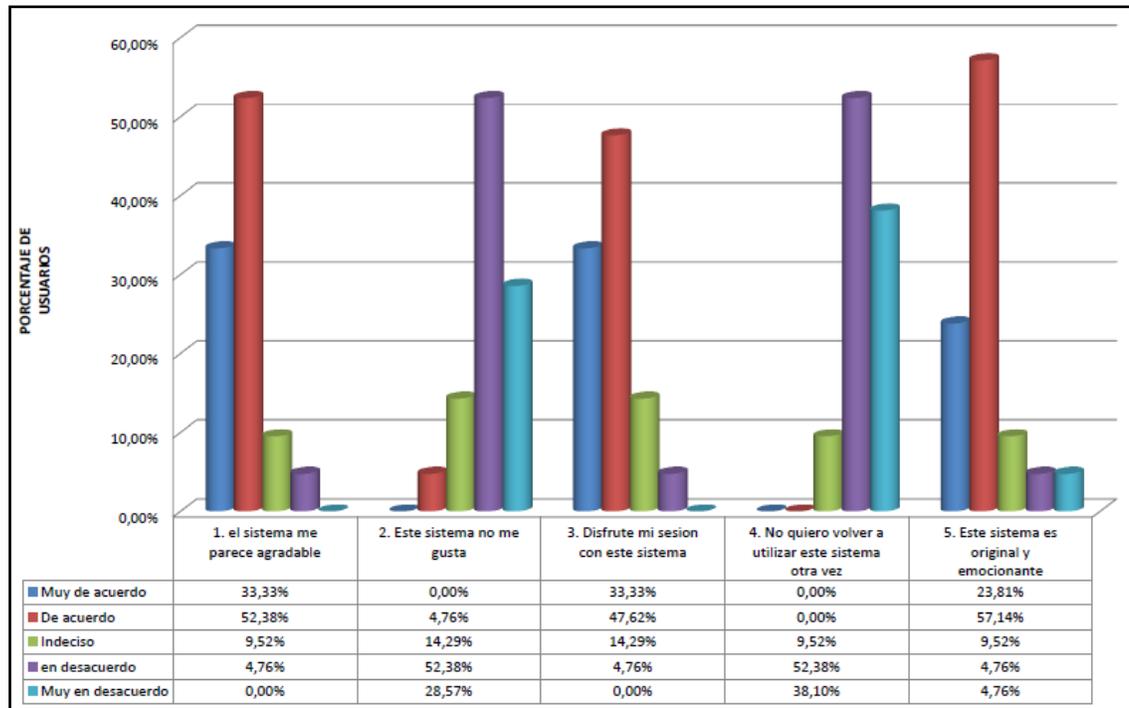


Figura 18. Atractivo del Sistema Controlado por Control Remoto.

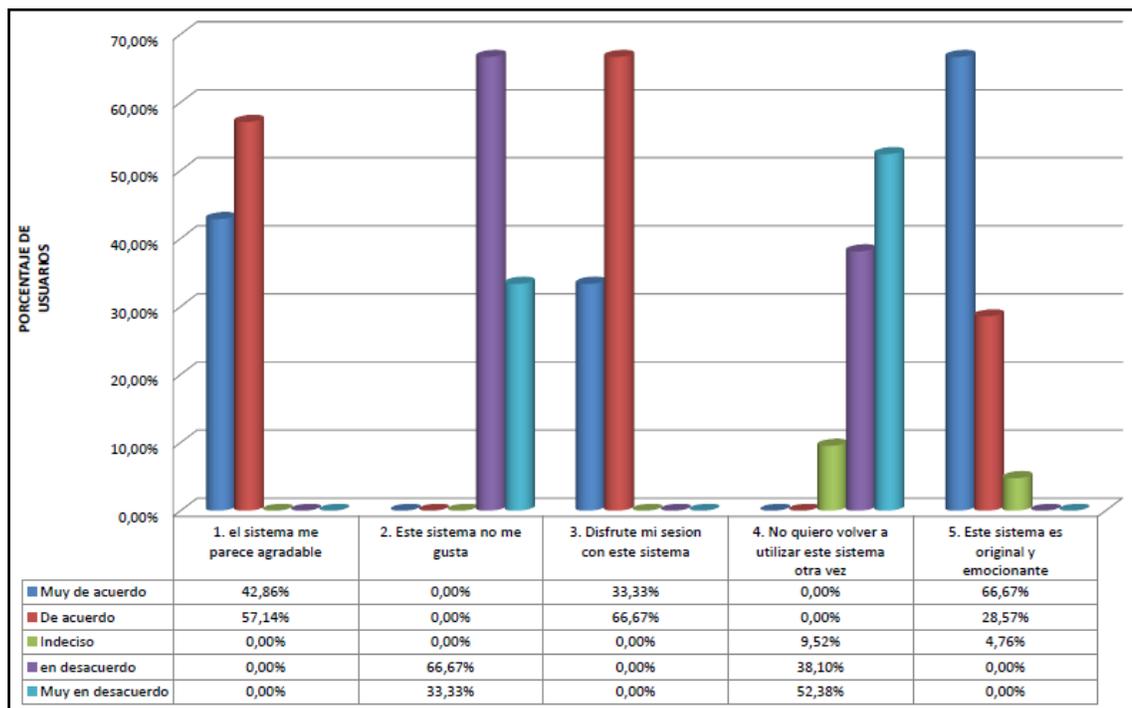


Figura 19. Atractivo del Sistema Controlado por Voz.

En general, en la figura 18, se observa que el sistema controlado por medio del control remoto posee una aceptación generalizada alrededor del 80% en todos los enunciados, presentando unos pocos usuarios indecisos, o inconformes acerca de algunos aspectos. Se resalta que un poco más del 90% de los encuestados les gustaría utilizar el sistema en otra oportunidad.

La figura 19, muestra claramente que la percepción del atractivo de la solución por voz es superior a la del sistema manejado por control remoto en todos los aspectos, teniendo a favor a más del 90% de los encuestados y con un nivel de duda mínimo.

En las preguntas abiertas, algunos usuarios reportaron que la interfaz gráfica de la solución es amigable, simple y de fácil navegación. Así mismo, ellos consideran que entre los dos sistemas elegirían el controlado por voz, dado que brinda comodidad y requiere poco esfuerzo.

4.3.3.2 Grado de Control

Las figuras 20 y 21, muestran los resultados obtenidos luego de la aplicación del CPS, en referencia al grado de control que percibe el televidente entre el sistema controlado por control remoto al controlado por voz, respectivamente.

En general, los resultados de las figuras 20 y 21, muestran la superioridad de la solución por voz en las afirmaciones 7 y 8 relacionadas a la facilidad de uso, resultando en una buena experiencia interactiva cuando el usuario logra cumplir con sus objetivos. En dichas afirmaciones hay una favorabilidad superior al 80%, existiendo pocos usuarios indecisos o insatisfechos con el grado de control.

Los resultados del control remoto para las expresiones 7 y 8, figura 20, describen una favorabilidad alrededor del 70%, con un porcentaje mayor de indecisos y usuarios insatisfechos en relación con el control por voz. Estos resultados se explican, de acuerdo a las preguntas abiertas, por el mayor control que ofrece la interacción por voz, minimizando el número de pasos para ejecutar una tarea, permitiendo acceder más fácilmente a la opción deseada mediante la pronunciación del comando sugerido por el sistema.

La afirmación 6 en relación con la interacción con comandos de voz, muestra un nivel de insatisfacción, la cual se nota en la división de opiniones entre las opciones, de acuerdo, indeciso, y en desacuerdo. En el sistema controlado por control remoto se obtienen mejores resultados, con una favorabilidad por encima del 57% en relación al 39% de insatisfacción. El bajo desempeño relacionado con la solución por voz, se asocia a la ejecución de comandos por error cuando el usuario habla, a la interferencia en la entrega de los comando, causados por el ruido del entorno y la duda acerca de los comandos que el sistema está en posibilidad de reconocer, problemas reportados en las preguntas abiertas.

Análisis del Uso de Interactividad basada en Voz para el manejo del Servicio de Video bajo Demanda

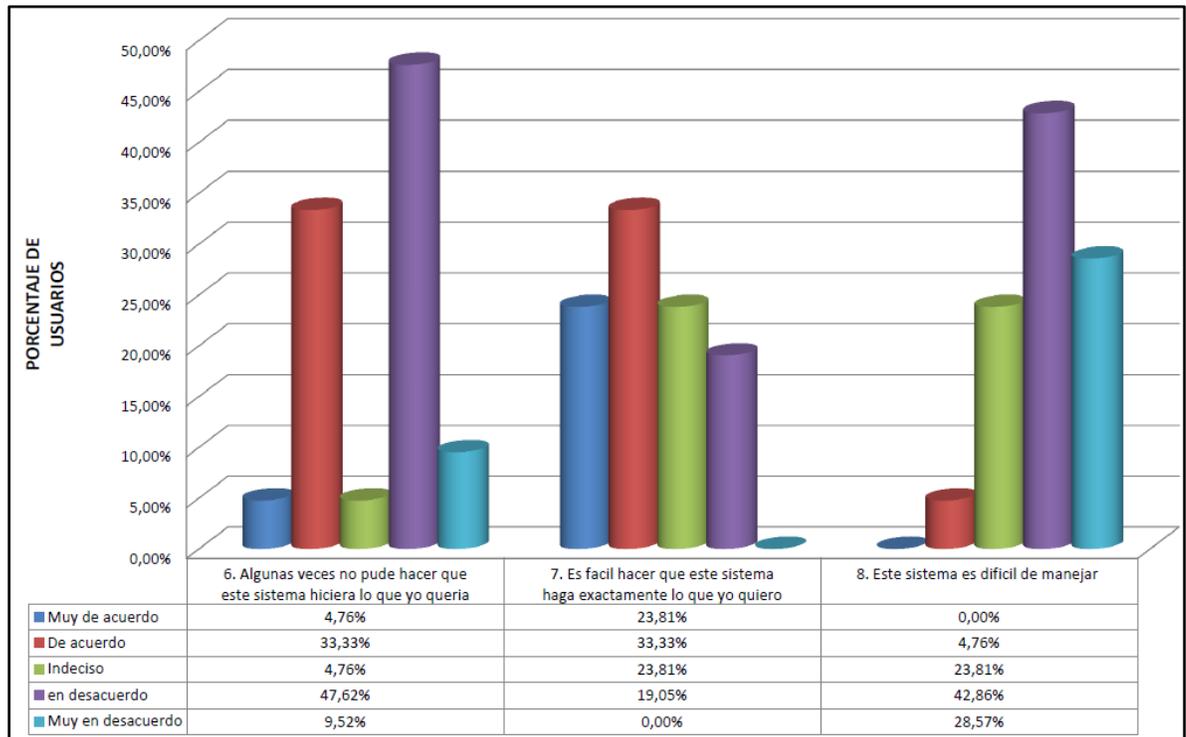


Figura 20. Grado de Control del Sistema con Control Remoto.

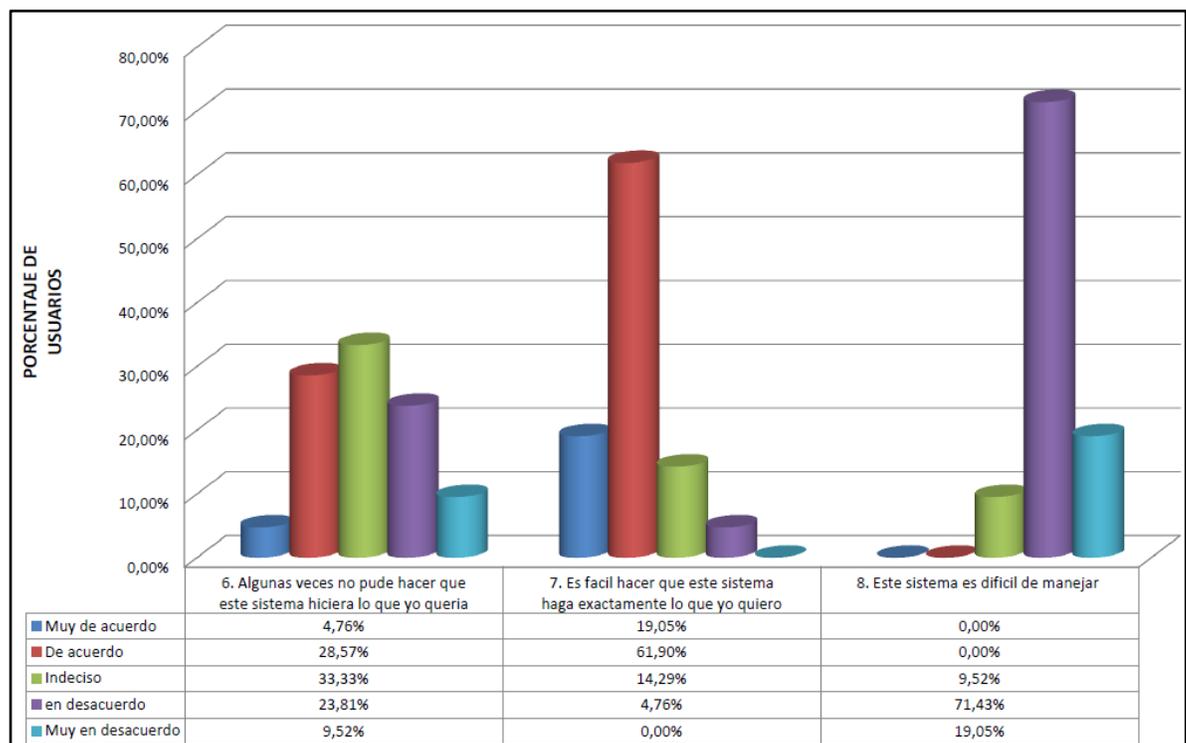


Figura 21. Grado de Control del Sistema con control por Voz.

De acuerdo a las preguntas abiertas, anexo B, los usuarios sugieren disponer de guías de interacción más claras y ampliar el número de comandos disponibles. Así mismo, se deben emplear micrófonos con mayor nivel de filtro de ruido del entorno.

4.3.3.3 Eficiencia

Las figuras 22 y 23, presentan los resultados en relación a la eficiencia del sistema, obtenidos de la aplicación del CPS.

De las figuras 22 y 23, se destaca la mediana superioridad de la voz en las afirmaciones 9 y 10, en contraste con la deficiencia manifestada en las afirmaciones 11 y 12. Los resultados de la expresión 9, muestran que para el 76.19% de los usuarios el sistema controlado por voz no requiere la ejecución de tareas aburridoras y repetitivas. Por tanto, la voz se considera una alternativa de interactividad un poco más eficiente y agradable frente al control remoto, que presenta un porcentaje de satisfacción del 66,67%.

La afirmación 10, relacionada con la precisión en la respuesta de los sistemas evaluados, indica que el servicio manejado por medio del control remoto cuenta con una favorabilidad en el 76,19% de la población, porcentaje muy cercado al 80,95% de la solución por voz. De esta manera, se considera que la propuesta del presente proyecto de investigación, cuenta con un alto grado de precisión, incluso mayor al del control remoto. Algunos usuarios reportaron que la interfaz por voz era intuitiva y lo guiaba durante el proceso de interacción, lo que facilitaba y aumentaba la precisión en el reconocimiento de los comandos adecuados.

La afirmación referente a la velocidad de respuesta del sistema, figuras 22 y 23, muestra un nivel de satisfacción superior del 70% para la interacción por control remoto, frente al 52,38% del sistema controlado por voz. Estos resultados se corroboran con las preguntas abiertas, en las cuales se reporta la lenta respuesta del sistema como el mayor problema de la solución.

La afirmación 12, permite observar el sistema manejado por voz presenta mayor número de errores repentinos en relación con el sistema con control remoto. Además del ruido y los problemas de reconocimiento reportados por los usuarios en las preguntas abiertas, algunas veces se presentaron problemas con el motor de reconocimiento por la caída del mismo.

El nivel de concentración percibido por los usuarios, expresión 13, se encuentra alrededor del 80% de favorabilidad entre la población. Sin embargo se observa que la interfaz por voz, cuenta con mejores resultados que los del control remoto. Este resultado se confirma con la simplicidad y diseño intuitivo de las interfaces gráficas y la interfaz por voz, reportados en las preguntas abiertas, anexo B.

Análisis del Uso de Interactividad basada en Voz para el manejo del Servicio de Video bajo Demanda

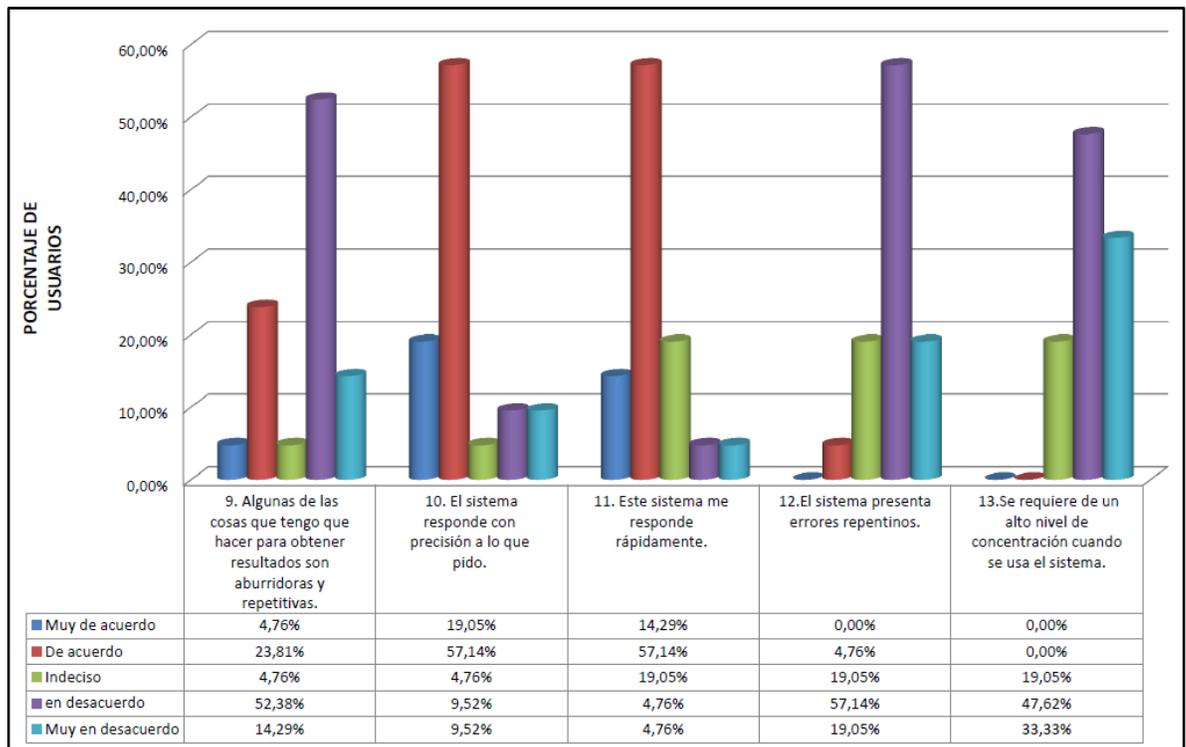


Figura 22. Eficiencia del sistema con Control Remoto.

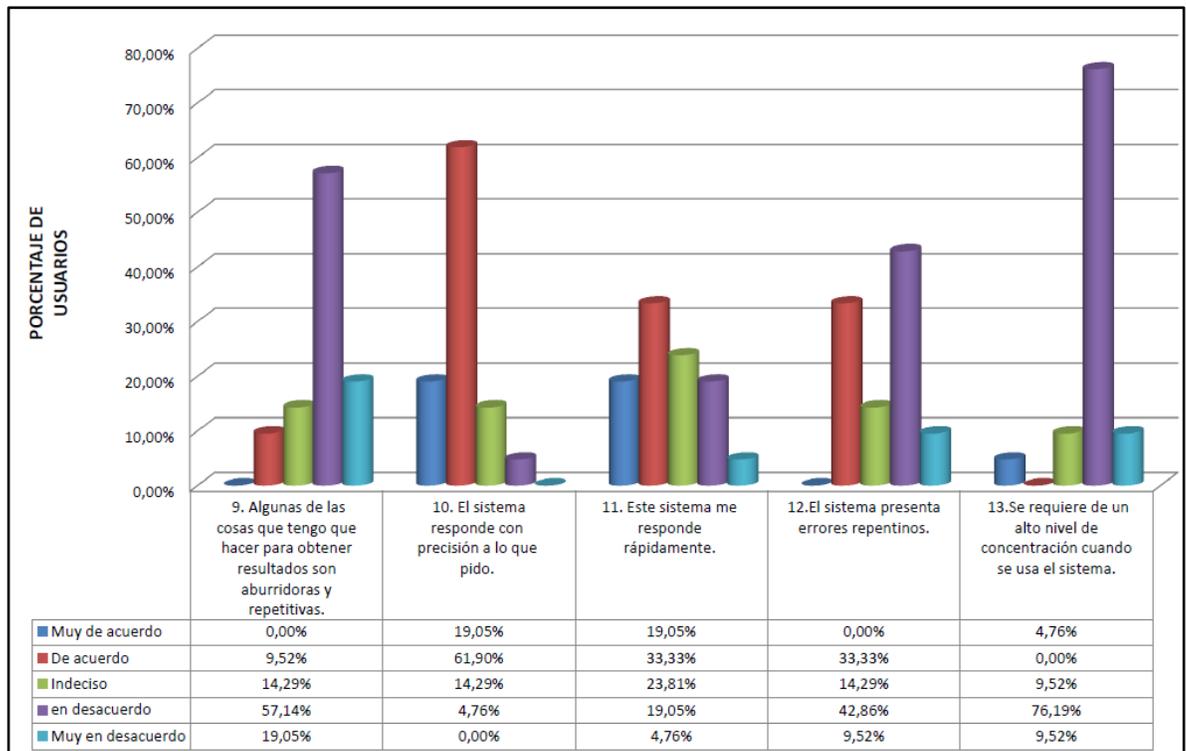


Figura 23. Eficiencia del Sistema con control por Voz.

4.3.3.4 Utilidad

Las figuras 24 y 25, muestran el bosquejo de las respuestas en relación con la utilidad de las soluciones para el control remoto y la voz, respectivamente. En éstas, se observa que la utilidad percibida de la solución de interactividad por voz, es superior a la del control remoto.

Los resultados de la afirmación 14, indican que la mayoría de los usuarios encuestados están de acuerdo o muy de acuerdo, en que la guía presente en la GUI de la solución por voz, les facilita encontrar lo que desean, ya que en la interfaz se les indica que decir. A partir de las preguntas abiertas se encontró que algunos usuarios creen que la interfaz por voz facilita el uso del servicio y les da comodidad porque no tienen que buscar el control remoto o no tienen que buscar el botón adecuado.

Con relación al enunciado 15, se observa que la voz es una interfaz más intuitiva que el control remoto, ya que un gran porcentaje de la población conocía acerca de los comandos que debía entregar. Sin embargo, con base en las observaciones y las preguntas abiertas, se determina que se debe mejorar la guía presente en las interfaces gráficas, e incrementar los comandos que el sistema debe reconocer.

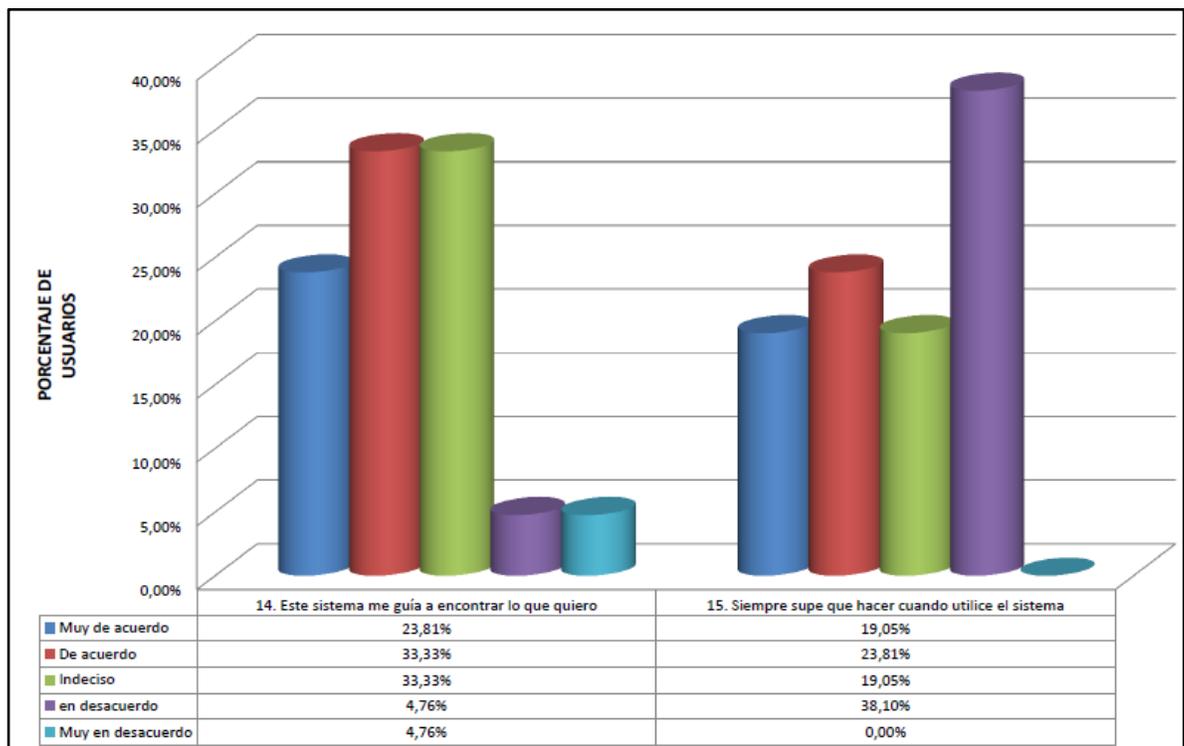


Figura 24. Utilidad Percibida del Sistema manejado con Control Remoto.

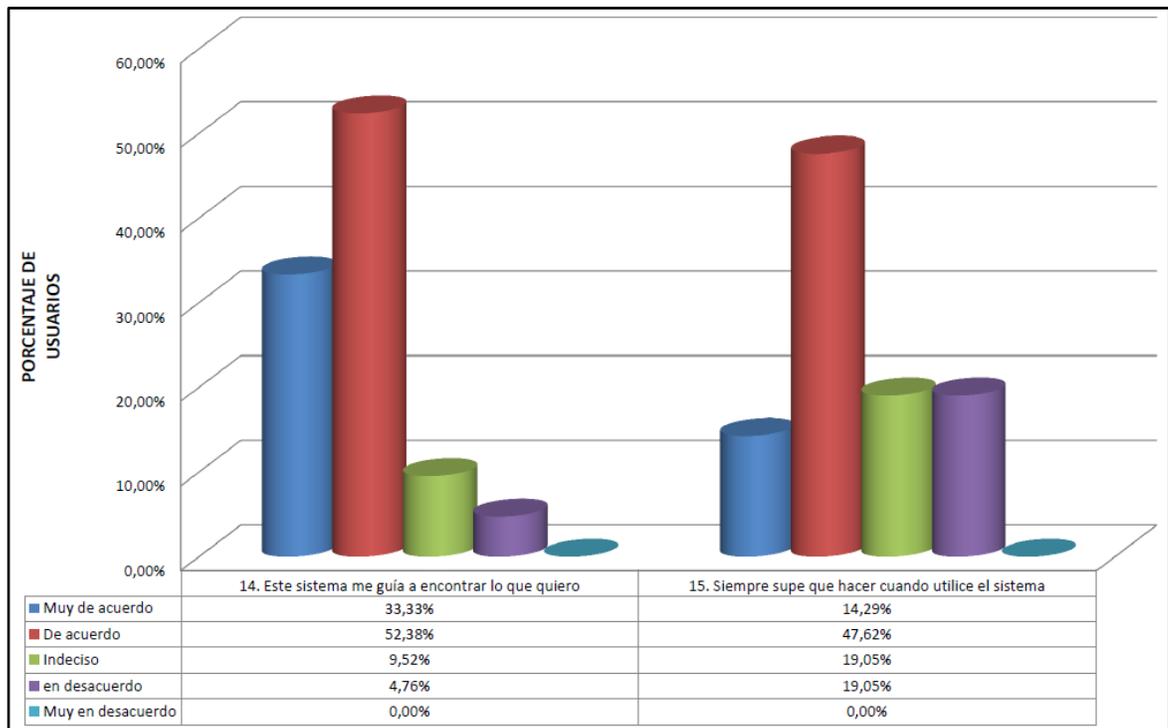


Figura 25. Utilidad Percibida del Sistema manejado por Voz.

4.3.3.5 Facilidad de Aprendizaje

El último aspecto a analizar desde la percepción del televidente, es la facilidad de aprendizaje. Comparando los resultados de las figuras 26 y 27, se observa que la solución controlada por voz, cuenta con resultados superiores en la mayoría de los enunciados.

Los resultados de la afirmación 16, indican que gran parte de los usuarios opinan que los sistemas motivaban a su uso, sin embargo, la solución por voz presenta mayor interés, lo cual se debe a lo novedoso del servicio de VoD para los televidentes, reportados en las preguntas abiertas, anexo B.

Con los resultados del enunciado 17, se observa que la mayor parte de los usuarios consideran que los sistemas son fáciles de usar, de manera tal, que si en otra ocasión hacen uso del mismo sistema no tendrán mayor problema en interactuar con él. No obstante, la interfaz por voz presenta un comportamiento más favorable, lo que se debe a que la interfaz es natural, intuitiva y le indica al usuario acerca del comando que debe pronunciar, según lo reportado por los usuarios.

Análisis del Uso de Interactividad basada en Voz para el manejo del Servicio de Video bajo Demanda

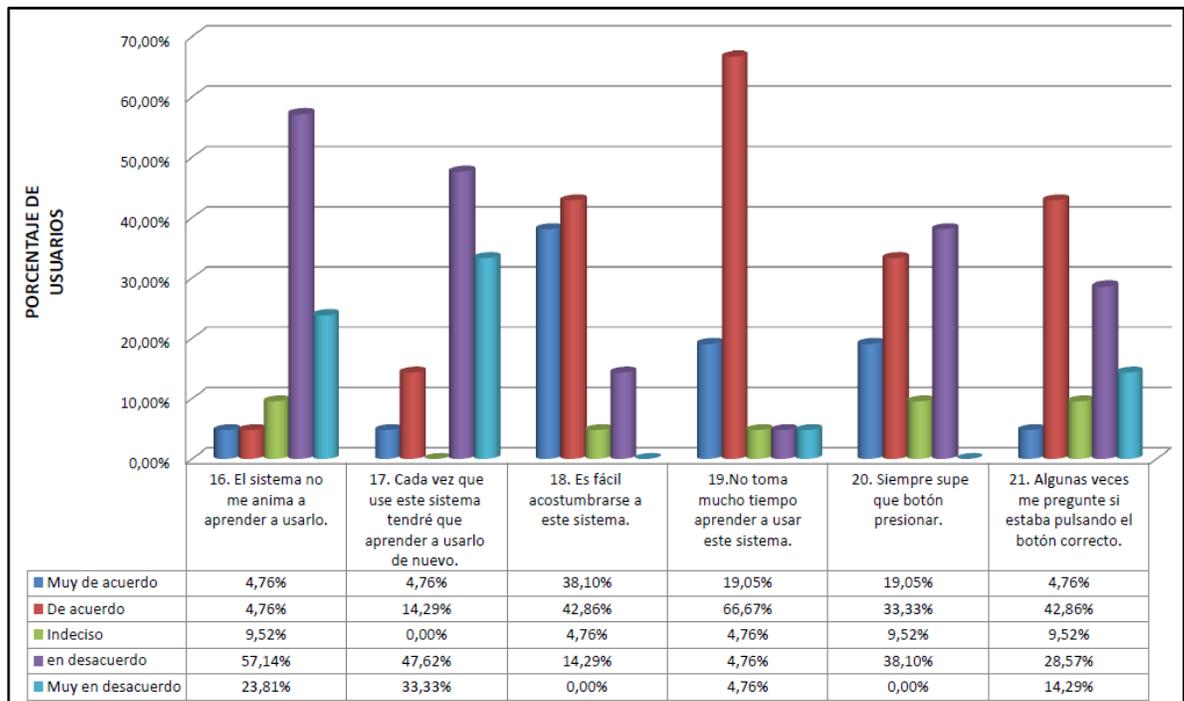


Figura 26. Facilidad de Aprendizaje del Sistema manejado por Control Remoto.

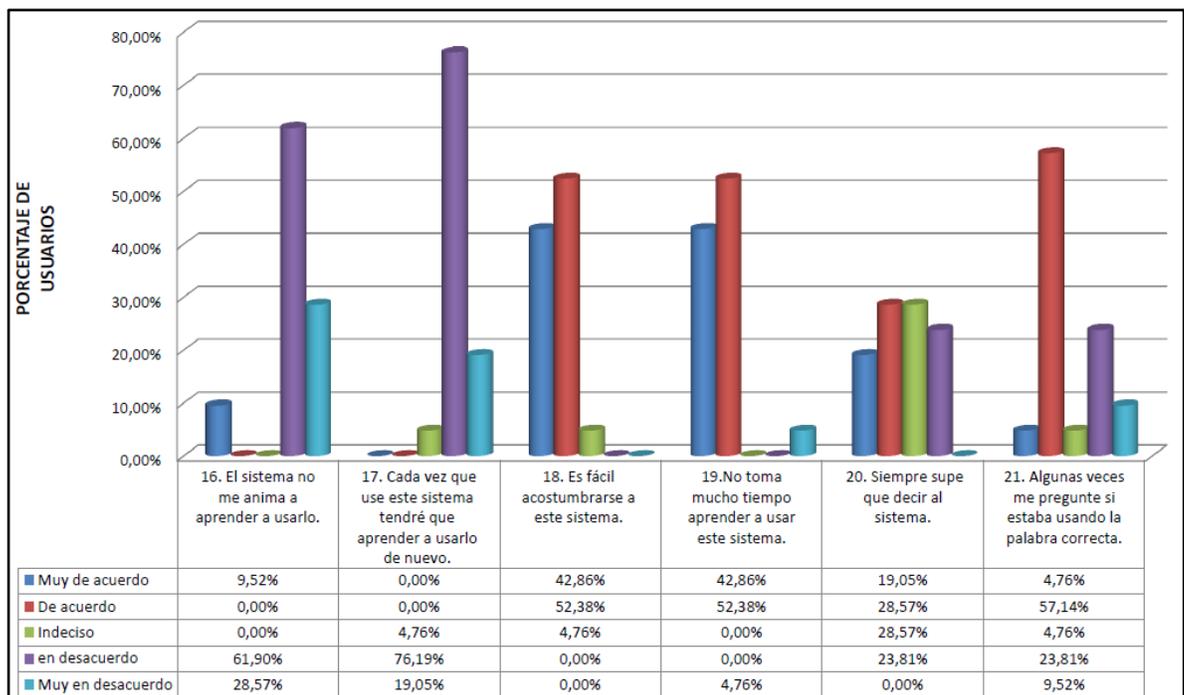


Figura 27. Facilidad de Aprendizaje del Sistema Manejado por Voz.

Los resultados del enunciado 18, señalan que un gran porcentaje de usuarios, 80% para el control remoto y 95% para la solución por voz, pueden acostumbrarse fácilmente a los

sistemas, dado que los dos cuentan con un diseño adecuado y posibilitan una experiencia agradable.

En cuanto al tiempo de aprendizaje, expresión 19, se observa que los dos sistemas muestran buenos resultados. La solución propuesta se apoya por un mayor porcentaje de población, debido a que la interfaz gráfica guía al usuario, evitando la carga cognitiva y el aprendizaje de comandos, tal como lo reportan las preguntas abiertas del CPS, anexo B. El servicio de VoD emulado con el DVD, conlleva a mayor tiempo de aprendizaje, ya que los televidentes deben encontrar el botón adecuado a presionar.

Los dos últimos enunciados presentan menor calificación para la solución por voz, como consecuencia, de la poca o nula experiencia previa con un sistema de este tipo. Además, a partir de la observación, se obtuvo que un gran porcentaje de la población, deseaba interactuar inmediatamente con la solución, sin la intención de leer las instrucciones presentes en la ayuda del sistema. Sin embargo, en las preguntas abiertas algunos usuarios piden mayor ayuda en la interfaz gráfica.

El análisis de la facilidad de aprendizaje de los sistemas, concluye el estudio de los resultados obtenidos, luego de la aplicación de los CPS al sistema emulado con el DVD y la solución propuesta en el presente proyecto de investigación.

Con el objetivo de evaluar la solución de interactividad basada en VoIP, desde la perspectiva de los elementos de red y algunos parámetros como el retardo, que podrían haber afectado el aspecto subjetivo, a continuación, se analizan los resultados obtenidos a partir del monitoreo de la red y los elementos involucrados en el funcionamiento de la solución.

4.4 EVALUACIÓN OBJETIVA

En esta sección se analiza los resultados de la evaluación objetiva realizada a la solución, en la cual, se desea determinar el desempeño a nivel de red, del control a través de VoIP para un servicio de VoD. Para ello se realizaron algunas capturas durante la realización del test de usuarios.

4.4.1 Comportamiento General de la Solución

Con el escenario planteado en la figura 14, se realiza la evaluación de la solución desde la perspectiva de red. Las figuras 28 y 29, describe el comportamiento del tráfico de red, de acuerdo a la cantidad de bytes en función del tiempo generados durante el uso de la solución por parte de uno de los usuarios.

Dicho grafico resultado de la captura de tráfico de una de las interacciones por medio del analizador de protocolos Wireshark. El escenario probado se trabajó con la red de datos

IP de la FIET, utilizando conexiones Ethernet 100BaseTX, los elementos descritos en el escenario (servidor, portátiles) se conectaron a la red por medio de Switches.

Los contenidos están codificados H.264, con una resolución de 720x480 con un entrelazado de 29,97 fps, que corresponden a un contenido con calidad estándar.

El códec de voz es el G.711 según lo planteado en la sección 3.5.2.1, es el más apropiado para un proyecto de esta naturaleza.

No fue posible la realización de una prueba de carga (con varios usuarios en un mismo instante), dadas las restricciones de la licencia del motor de reconocimiento que solo reconoce un usuario por sesión (un solo puerto).

La transferencia de información inicia cuando se establece la llamada a la extensión en el servidor de comunicaciones IP (Asterisk) y desde el navegador se envía una orden HTTP (GET) al servidor WEB a la primera interfaz de la solución. En la gráficas 24 y 25 se observa que una vez se inicia la llamada (línea azul), empieza el proceso de interacción. Una vez el usuario ha dado un comando de voz y este ha sido reconocido de forma exitosa por el motor de reconocimiento, se envía el comando reconocido por medio de TCP (puntos en verde) al Applet, la ejecución de dichos comandos en el lado del cliente puede conllevar a que se den peticiones GET de otros elementos de la interfaz al servidor WEB (línea negra), que se dé inicio a una petición (PLAY, PAUSE, TEARDOWN) al servidor de Streaming (comportamiento de la línea roja) o que se dé la ejecución de un comando a nivel local en el navegador.

En la figura 29, también se puede observar el tráfico generado por Asterisk (barras en morado), dicho tráfico es a ráfagas y se genera como indicador al usuario de que el sistema ha realizado un proceso de reconocimiento y está listo para reconocer un nuevo comando. De la gráfica se abstrae que en general dicho trafico está muy asociado al envío de los mensajes TCP, a excepción de algunos casos en los que no hay envío de mensajes TCP, causado o por que el comando reconocido no paso un nivel de umbral establecido en la programación o a que el usuario ha desactivado el mecanismo de interactividad y el sistema está a la espera del comando de activación.

De los distintos tráficos que se manejan en la solución (VoIP, Video, TCP, entre otros), los que afectan el desempeño del mecanismo de interacción basado en voz, son el de VoIP y el de TCP. Sin embargo se ha analizado cual es el comportamiento de flujo de video en búsqueda de ver como se dio la entrega de los contenidos.

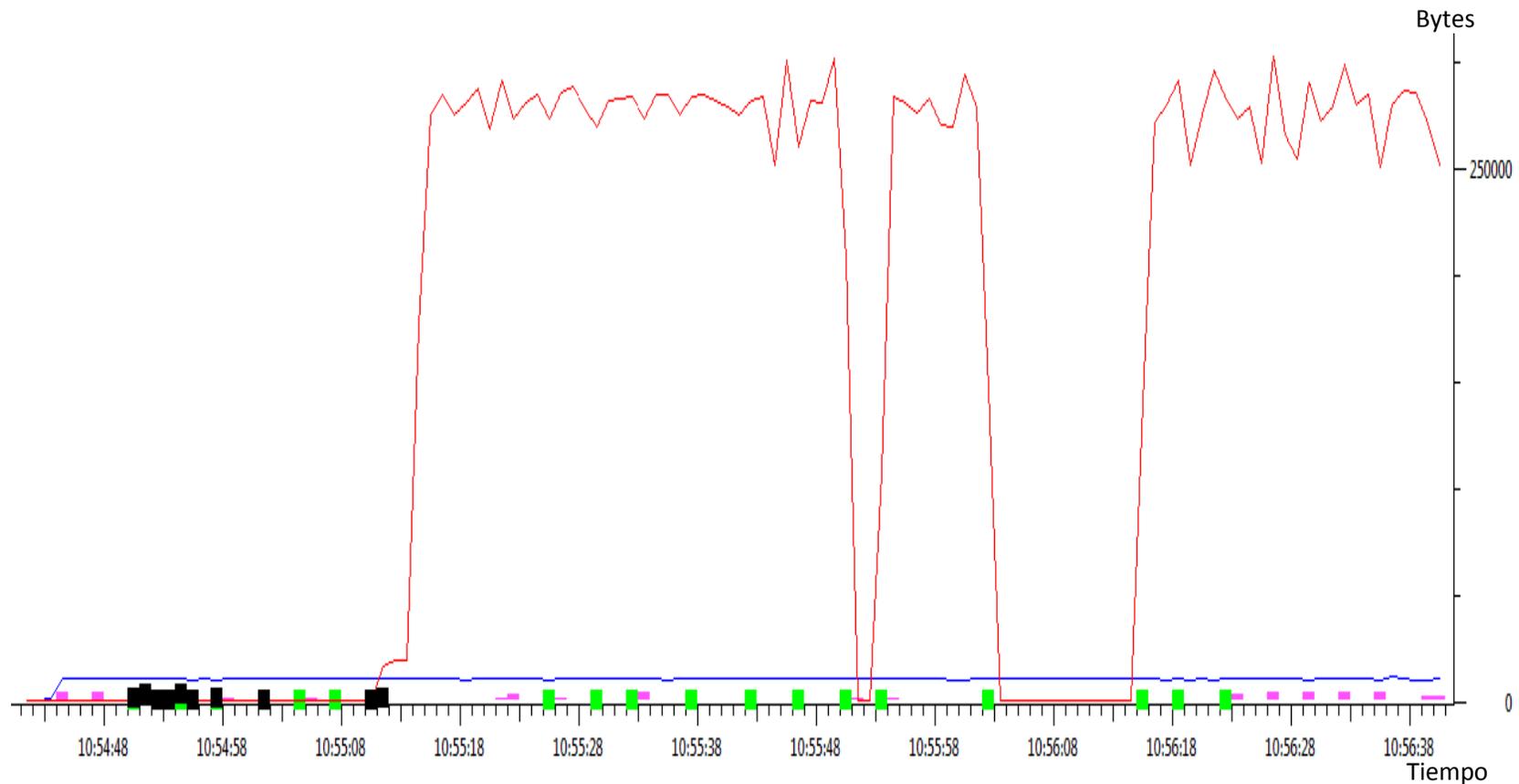


Figura 28. Tráfico total de la solución piloto.

- Tráfico HTTP.
- Tráfico de video.
- Tráfico de voz con origen en el cliente.
- Tráfico de voz con origen en el servidor.
- Mensajes TCP que envían los comandos reconocidos hacia el cliente.

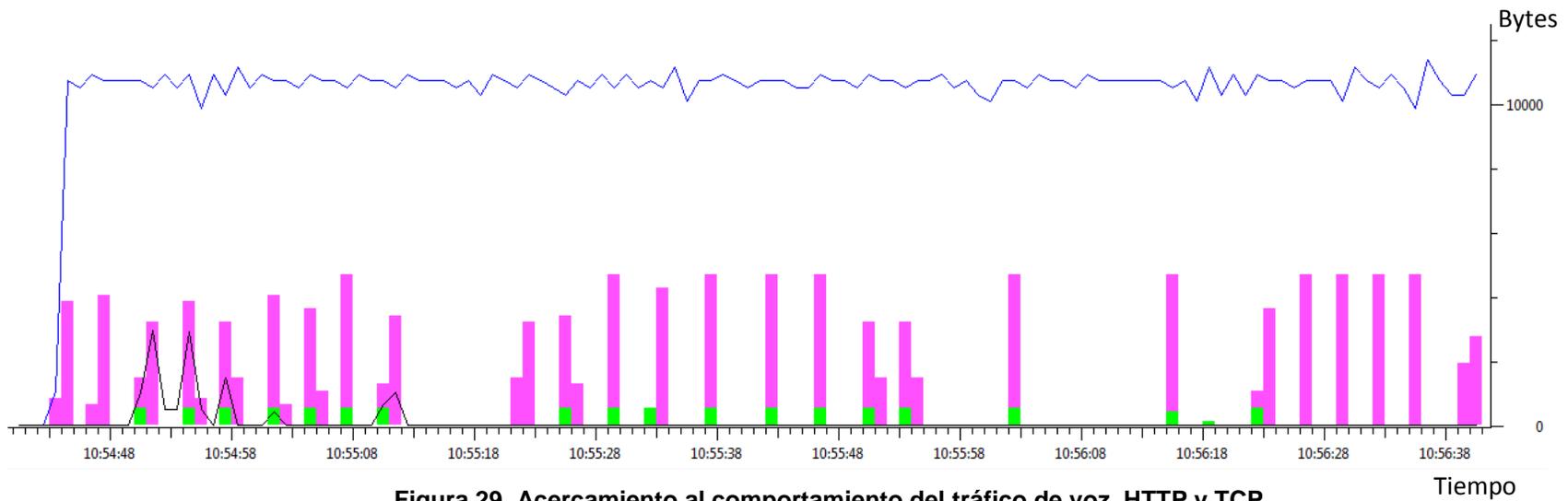


Figura 29. Acercamiento al comportamiento del tráfico de voz, HTTP y TCP.

- Tráfico HTTP.
- Tráfico de video.
- Tráfico de voz con origen en el cliente.
- Tráfico de voz con origen en el servidor.
- Mensajes TCP que envían los comandos reconocidos hacia el cliente.

4.4.2 Comportamiento del Flujo de Video

De la figura 28, se observa que la tasa de transferencia, varía entre 2000 kbps y 2400 kbps, valores que corresponden con la Recomendación de Mínima Velocidad Binaria para un Canal SDTV para Contenidos de VoD [61].

Durante el uso de la solución de interactividad por voz, se observan pequeños retardos en el inicio de la reproducción del contenido multimedia, éste se debe al retardo y jitter del transporte de los flujos de video desde el servidor de video streaming hacia el cliente multimedia. La tabla 21, contiene los valores promedio de retardo y jitter, obtenidos por medio de OmniPeek a partir de la captura de paquetes realizada a través de WireShark.

Dado que el comportamiento de la solución es estable e independiente entre cada interacción, se analiza el comportamiento de una de las interacciones realizadas.

Tabla 21. Comportamiento del Flujo de Video.

Flujo de Información	Parámetro		
	Jitter	Retardo	Pérdida de paquetes
Video (H.264)	31,591ms	58 ms	0

De la tabla 21, se observa que el retardo del video es menor a 200 ms y el jitter es menor a 50 ms, valores máximos establecidos para la transmisión de contenidos en tiempo real para un servicio de VoD [85]. De esta manera, es posible que el retardo en la reproducción del contenido multimedia, sea consecuencia de la complejidad de decodificación del códec H.264, además del tiempo que requiere el buffer de compensación de jitter en el terminal de usuario.

Cabe resaltar el buen desempeño de la red en relación con la nula pérdida de paquetes. El anexo C, contiene las capturas en pantalla de las gráficas entregadas por OmniPeek.

4.4.3 Comportamiento de la Comunicación VoIP

Teniendo en cuenta la información de la figuras 29 y la figura 30 (obtenida mediante CacePilot), se evidencia un ancho de banda consumido alrededor de 90 kbps, valor que es superior a los valores teóricos de 64Kbps que ofrece el códec G711, lo que se debe a la adición de las diferentes cabeceras RTP, UDP, IP y ETHERNET. Aunque este valor es grande en comparación con los demás Codecs [86] vale la pena usar este códec en la medida que los errores de reconocimiento pueden verse disminuidos.



Figura 30. Ancho de banda consumido por el códec G.711.

Para analizar el desempeño de la comunicación establecida a través de VoIP, se introduce en Cace Pilot, los paquetes capturados por medio de WireShark. Dicho programa entrega los valores consignados en la tabla 22, para el jitter, retardo y pérdida de paquetes promedio de la comunicación establecida durante el uso de la solución propuesta.

Tabla 22. Comportamiento de la Comunicación de VoIP.

Código	Parámetro		
	Jitter	Retardo	Pérdida de Paquetes
G.711	8,453 ms	57 ms	0,4469%

De acuerdo a los valores presentes en la tabla 22, se determina que el desempeño de la comunicación VoIP es bueno, debido a que la pérdida de paquetes es mínima y los valores de retardo y el Jitter se encuentran por debajo del máximo recomendado [87], [88] que son de 150 ms para el retardo y 100 ms para el jitter.

El anexo C, contiene las capturas en pantalla de los resultados entregados por Cace Pilot consignados en la tabla 22.

4.4.4 Comportamiento Socket TCP

El comportamiento de la conexión, a través de la cual, se envía los comandos del televidente desde la interfaz en el módulo de reconocimiento del habla hacia el cliente multimedia, siguen el diagrama de flujo mostrado en la figura 31.

El comportamiento de los paquetes TCP, se obtiene a partir del análisis de 30 transmisiones que corresponden al envío de 30 comandos desde el módulo de reconocimiento del habla hacia el cliente multimedia. De acuerdo a lo anterior, se estima que la duración del diagrama de flujo, mostrado en la figura 31, es de 10 ms. En el Anexo C, se describe la forma en la cual se mide dicho tiempo.

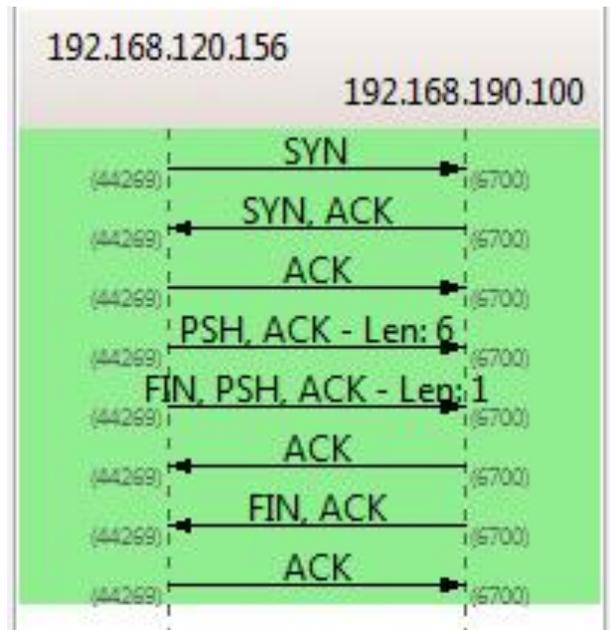


Figura 31. Diagrama de Flujo de los paquetes TCP.

Dado el buen desempeño de la red de comunicaciones, no se presentaron errores de transmisión, ni retransmisiones.

4.4.5 Retardo Total

De acuerdo al proceso de interacción descrito en el capítulo 3 y mostrado en la figura 11, se procede a analizar el retardo generado en cada elemento de la solución, y de esta manera, encontrar el retardo total de la solución, desde el momento en el cual el cliente VoIP envía la voz para su reconocimiento, hasta el momento en el cual lo recibe el Applet:

1. En concordancia con [89], el retardo de codificación/decodificación para G.711 se encuentra alrededor de 0,125 ms.
2. El retardo transmisión de la voz es alrededor de 57 ms, tabla 22.
3. El retardo de procesamiento causado por el motor de reconocimiento se estima en 1260,138 ms, el cálculo de este valor se explica en el anexo C.
4. El retardo de transmisión del mensaje TCP, de acuerdo a 4.2.3 es del orden 10 ms.

La suma de los valores de retardo antes mencionados, es el retardo total promedio de la solución, que se estima en 1327,263 ms. Por tanto, se observa que el motor de reconocimiento, es el causante de gran parte del retardo durante el uso de la solución.

Debido a que el retardo total de la solución es superior a (1) segundo [89], se justifica el porqué de la respuesta lenta, reportada por los usuarios en las preguntas abiertas, así como, el resultado de la baja conformidad con la respuesta del sistema de la figura 23, anexo B.

Con el análisis del retardo total, se concluye el análisis de la solución piloto de interactividad basada en voz desde el aspecto objetivo de la evaluación. Así mismo, se termina con el proceso de evaluación de la solución propuesta en el capítulo 3.

A lo largo de este capítulo se ha explicado el proceso de evaluación de la solución propuesta, desde el aspecto subjetivo de los usuarios, y el aspecto objetivo de la comunicación de red. Posteriormente, los resultados obtenidos, han sido analizados. Con base en dichos resultados y el estudio teórico realizado en los capítulos previos, a continuación se presentan las conclusiones y trabajos futuros derivados del presente proyecto de investigación.

Capítulo 5. CONCLUSIONES, RECOMENDACIONES Y TRABAJO FUTURO

5.1 CONCLUSIONES

- ✓ Actualmente existen muchos mecanismos de interactividad disponibles en el ambiente de los servicios de entretenimiento, cada uno de los cuales posee sus ventajas y desventajas. La selección de uno de ellos en particular dependerá de la población objetivo que se quiera atender y de las condiciones y limitaciones del ambiente donde se utilicen (sección 1.3).
- ✓ Los mecanismos de interactividad basados en el control por voz, se tornan como una opción potencial para los usuarios de servicios interactivos, tal como lo demuestran los productos de Amulet Devices®, que tienen una solución para los sistemas Windows Media Center®.
- ✓ Al plantear una solución óptima para el control por voz de servicios interactivos es necesario contar con dispositivos de alta capacidad de procesamiento, para que atiendan las solicitudes de los usuarios (motor de reconocimiento de voz, procesamiento de la petición, etc.) en tiempos cortos y puedan procesar múltiples solicitudes simultáneamente (sección 3.5 y 3.6).
- ✓ La evaluación de una solución de interactividad basada en voz puede realizarse desde dos perspectivas: la objetiva y la subjetiva. La objetiva evaluará el comportamiento técnico de la interacción, esto es, los tiempos de retardo a las solicitudes, demora en el reconocimiento de la voz, jitter y pérdida de paquetes. La subjetiva evaluará la percepción que tienen los usuarios al utilizar el sistema e interactuar con el mismo a través de la guía de programación electrónica, y conocer si responde adecuadamente a sus expectativas. Las dos perspectivas de evaluación de forma conjunta pueden dar una apreciación coherente del comportamiento de la solución de interactividad basada en voz (capítulo 4).
- ✓ Acorde a lo evaluado en el prototipo de solución de interactividad basada en voz, es necesario que este tipo de solución se cree para ambientes bajos de ruido, donde el número de personas sea reducido y donde las condiciones de

aislamiento de ambientes externos (calles, ambientes industriales, ambientes comerciales) sea alto; por tal razón una solución de este tipo es más adecuada para las salas y habitaciones de los hogares y habitaciones de hotel (sección 3.6).

- ✓ El prototipo de solución propuesto para evaluar la interactividad basada en voz, se desarrolló en un salón de trabajo de grado de la universidad, y contó con la participación en el proceso de evaluación subjetiva con personas (21) de diversas edades y variados niveles de educación, por lo cual en este contexto y acorde a las respuestas obtenidas por los participantes, se identifica a la solución de interactividad basada en voz como interesante y en muchos de los ítems evaluados más aceptable que el control remoto (sección 4.3.3).

5.2 RECOMENDACIONES

- ✓ Al proponer soluciones de interactividad basada en voz, se debe procurar disponer de motores de reconocimiento de alta calidad y a su vez de sistemas de procesamiento de alta capacidad.
- ✓ Se recomienda que antes de evaluar una solución de interactividad basada en voz, se realice con las personas que van a participar en el proceso de evaluación subjetiva una corta capacitación e ilustración sobre la interfaz gráfica de usuario que van a utilizar y sobre la forma en que el servicio busca operar. A su vez, es necesario generar una guía o folleto donde se ilustren los pasos que se deben seguir en la interacción con el sistema para que la puedan revisar por si olvidan algo.
- ✓ Luego de finalizar la fase de evaluación de la solución, se estima que la solución, con las características descritas en 4.2, puede soportar un número promedio entre 15 y 20 usuarios, dado que los servidores de video, web, de comunicaciones IP y motor de reconocimiento se encuentran configurados en un solo equipo. Por tanto, se recomienda hacer uso de equipos especializados y de alto rendimiento en un entorno real. Además, se recomienda considerar el uso de *Wowza Streaming Server*, como servidor de video, ya que éste podría incrementar el número de usuarios simultáneos que utilizan el servicio. *Wowza Streaming Server*, no se instaló dado que éste cuenta con una licencia privativa y no estaba contemplado dentro de los alcances del proyecto realizar pruebas de carga de los servidores.

5.3 TRABAJO FUTURO

- ✓ Reemplazar los dispositivos utilizados en este piloto de solución de interactividad basada en voz: Equipo de Computo y Softphone, servidor de video, motor de reconocimiento, unidad STB, etc., los cuales fueron fundamentalmente soportados en computadores normales, por equipo más especializado y evaluar los diversos tipos de codecs de voz que se podrían utilizar para la interactividad.

- ✓ Desarrollar un piloto de interactividad basada en voz, donde se integren sistemas de seguridad, procesos de tarificación, entre otros y se realicen pruebas de carga de solicitudes y tiempos de respuesta ante altos volúmenes de usuarios simultáneos.

- ✓ Desarrollar una interfaz gráfica de usuario para una solución de interactividad basada en voz desde los principios de usabilidad y evaluarla desde dicha perspectiva.

REFERENCIAS

- [1] R. Eriksson and F. Sjögren, "Enhancing the User Experience with New Interaction Techniques for Interactive Television," M.S. thesis, Umeå Univ., Umeå, Sweden, 2007.
- [2] L. Miranda, E. Hayashi and M. Baranauskas, "An Experimental Scenario to Investigate the Remote Control as Artifact for Interaction with Television," Inst. Computação, Univ. Estadual Campinas, Campinas, Brasil, Tech. Rep. IC-09-22, 2009.
- [3] L. Miranda, L. Piccolo and M. Baranauskas, "Uma Proposta de Taxonomia e Recomendação de Utilização de Artefatos Físicos de Interação com a TVDI," in *Workshop Perspectives, Challenges and Opportunities Human-Computer Interaction Latin America (CLIHC)*, Rio de Janeiro, 2007.
- [4] O. Halset, "IP Television – Content on Demand: A usability and user experience evaluation of an IP television portal by the Norwegian broadcaster TV 2 Interaktiv," M.S. thesis, School Media Technol., Royal Inst. Technol., Stockholm, Sweden 2006.
- [5] T. Kunert, "Interactive TV Applications and Their Context of Use," in *User-Centered Interaction Design Patterns for Interactive Digital Television Applications* (Human-Computer Interaction Series), J. Karat and J. Vanderdonckt, Eds., New York: Springer, 2009, pp. 19-46.
- [6] R. Picard, "A Consumer Perspective on Digital Terrestrial and Interactive Television," in *Digital Terrestrial Television in Europe*, A. Brown and R. Picard, Eds. New Jersey: Lawrence Erlbaum Associates, 2005, pp.135-149.
- [7] D. Bacchan, F. Afonso, G. Araujo and R. Fortes, "Developing content for iTV: a case study," in *2nd European Conf. Interactive Television: Enhancing the Experience*, Brighton, United Kingdom, 2004, pp. 173-175.
- [8] K. Lu, "Interaction Design Principles for Interactive Television," M.S. thesis, Georgia Inst. Technol., Atlanta, GA, United States of America, 2005.

- [9] M. Estebanell, "Interactividad e interacción," *Revista Latinoamericana de Tecnología Educativa*, vol.1, no.1, pp. 11-18, 2002.
- [10] H. Benoit, *Digital Television: Satellite, Cable, Terrestrial, IPTV, Mobile TV in the DVB Framework*, J. Ward, Ed., 3rd ed., United States of America: Focal Press, 2008.
- [11] G. Held, "Television Concepts", in *Understanding IPTV*, Boca Raton, FL: Auerbach Publications, 2007, pp. 49-83.
- [12] P. Sotomayor, "Análisis de los Estándares de Televisión Digital Terrestre TDT y Pruebas de Campo Utilizando Equipos de Comprobación Técnica de la Superintendencia de Telecomunicaciones," M.S. thesis, Escuela Politécnica Nacional, Quito, Ecuador, 2009.
- [13] A. Rost, "Pero, ¿de qué hablamos cuando hablamos de Interactividad?," in Congr. ALAIC/IBERCOM, La Plata, Argentina, 2004.
- [14] J. F. Jensen, "'Interactivity': Tracking a New Concept in Media and Communication Studies," *Nordicom Rev.*, vol. 19, no. 1, pp. 185-204, 1998.
- [15] A. García, "Principios de interactividad: Televisión Interactiva y Realidad Virtual," *Enlaces*, vol. 7, jun., 2007.
- [16] H. Vivar, A. García, "La interactividad: concepto y factor de impulso de la TDT. Un caso práctico: Proyecto Sports ITV," *Sphera Pública*, vol. 9, pp. 207-222, 2009.
- [17] W. Godzic, "Various faces of interactivity: remarks on television," *ICONO 14: Revista de Comunicación y Nuevas Tecnologías*, no. 15, pp. 22-36, 2010.
- [18] L. A. Franch, "Los nuevos formatos de publicidad interactiva en television," in *Congreso Internacional Fundacional AE-IC*, Santiago de Compostela, 2008.
- [19] E. Prado, R. Franquet, F. X. Ribes, M. T. Soto and D. Fernández, "La publicidad televisiva ante el reto de la interactividad," *Questiones Publicitarias*, vol. 1, no. 12, pp. 13-28, 2007.
- [20] C. Rees, "BBC Enhanced TV Formats," British Broadcasting Corporation, London, United Kingdom, 2004.
- [21] L. Christensen, "The 'Impact' of Interactivity on Television Consumption," Ph.D. dissertation, Dublin Univ., Dublin, Ireland, 2002.
- [22] G. Hölbling, T. Rabl and H. Kosch, "Overview of Open Standards for Interactive

- TV (iTV),” *Multimedia Semantics - The Role of Metadata*, Feb., 2008.
- [23] A. Cobo, “Nuevas aplicaciones y servicios interactivos de TDT,” thesis, Departamento de Teoría de la Señal y Comunicaciones, Escola Tècnica Superior d’Enginyeria de Telecomunicacions de Barcelona, Universitat Politècnica de Catalunya, 2009.
- [24] A. Ibrahim, J. Lundberg y J. Johansson, “Speech Enhanced Remote Control for Media Terminal,” in *EUROSPEECH: 7th European Conf. Speech Commun. and Technol.*, Aalborg, Denmark, 2001.
- [25] W. Freeman, and C. Weissman, “Television Control by Hand Gestures,” Mitsubishi Elect. Res. Labs., Cambridge, MA, United States of America, Tech. Rep. TR94-24, 1994.
- [26] J. Oliveira and V. Martins, “Non-Functional Requeriments to Voice User Interface on Interactive Television: An Initial Study,” en *IADIS Int. Conf. Interfaces and Human Comput. Interaction*, Amsterdam, The Netherlands, 2008.
- [27] A. Berglund, “Future TV Interaction: Augmenting Existing Interaction Techniques”, Dept. Comput. and Inform. Sci., Linköping Univ., Linköping, Sweden, 2003.
- [28] N. Enns y S. MacKenzie, “Touchpad-Based Remote Control Devices,” in *ACM SIGCHI Conf. Human Factors Comp. Syst. "Making the Impossible Possible"*, Los Angeles, CA, United States of America, 1998.
- [29] Telefónica I+D, “Interfaces multimodales,” in *Las Telecomunicaciones y la Movilidad en la Sociedad de la Información*, División de Relaciones Corporativas y Comunicación de Telefónica I+D, Fundación Telefónica, 2005, pp. 289- 304.
- [30] K. Chorianopoulos and D. Spinellis, “User Interface Evaluation of Interactive TV: A Media Studies Perspective,” *Universal Access in the Information Society*, vol. 5, no. 2, pp. 209–218, Aug., 2006.
- [31] J. Calle, P. Martínez, D. del Valle and D. Cuadra, “Towards the Achievement of Natural Interaction,” *Engineering the User Interface*, vol. 1, no. 19, 2009.
- [32] S. O’Modhrain and I. Oakley, “Touch TV: Adding Feeling to Broadcast Media,” in *1st European Conf. Interactive Television: Viewers Actors*, Brighton, United Kingdom, 2003.
- [33] F. Lima, J. V. de Lima, R. A. de Nevado, “An Alternative for the Interaction with Digital TV,” in *Interactive Digital TV in Emergent Countries*, Tampere, Finland, 2010.

- [34] L. Miranda, H. Hornung and M. Baranauskas, "MuTIS: A Gesture Based Interaction Model for iDTV," Inst. Computação, Univ. Estadual Campinas, Campinas, Brasil, Tech. Rep. IC-09-004, 2009.
- [35] M. Bhuiyan and R. Picking, "Gesture-controlled user interfaces, what have we done and what's next?," in *5th Collaborative Res. Symp. Security, E-Learning, Internet and Networking (SEIN)*, Darmstadt, Germany, 2009, pp. 59-60.
- [36] A. Ibrahim, "Multimodal Interaction with Interactive TV," Dept. Comput. and Inform. Sci., Linköping Univ., Linköping, Sweden, 2000.
- [37] A. Ibrahim and P. Johansson, "Multimodal Dialogue Systems for Interactive TV Applications," in *4th IEEE Int. Conf. Multimodal Interfaces*, Pittsburgh, PA, United States of America, 2002, pp. 117 – 122.
- [38] S. Malik and J. Laszlo, "Visual Touchpad: A Two-handed Gestural Input Device," in *6th Int. Conf. Multimodal interfaces (ICMI)*, State College, PA United States of America, 2004, pp. 289 – 296.
- [39] R-D. Vatavu and S-G. Pentiu, "Interactive Coffee Tables: Interfacing TV within an Intuitive, Fun and Shared Experience," in *6th European Conf. Changing Television Environments*, Salzburg, Austria, 2008, pp. 183–187.
- [40] M. Nielsen, M. Störring, T. B. Moeslund and E. Granum, "A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for Man-Machine Interaction," Aalborg Univ., Aalborg, Denmark, Technical Report CVMT 03-01, 2003.
- [41] S. Ahn, T-S. Lee, I-J. Kim, Y-M Kwon and H-G.Kim, "Large display Interaction using video Avatar and Hand Gesture Recognition," in *Int. Conf. Image Analysis and Recognition (ICIAR)*, Porto, Portugal, 2004, pp. 261-268.
- [42] M. Chen, L. Mummert, P. Pillai, A. Hauptmann and R. Sukthankar, "Controlling your TV with Gestures," in *11th ACM SIGMM Int. Conf. Multimedia Inform. Retrieval*, Philadelphia, PA, United States of America, 2010, pp. 405-408.
- [43] J. Stegmann, K. Henke and R. Kirchherr, "Multimodal Interaction for Access to Media Content," presented at 12th Int. Conf. ICIN, Bordeaux, Francia, 2008.
- [44] Instituto Nacional de Tecnologías de la Comunicación, *Guía de Recomendaciones de Accesibilidad en TDT*, España: Instituto Nacional de Tecnologías de la Comunicación (INTECO), 2009.

- [45] J. Fails and D. Olsen Jr., "MagicWand," Comp. Sci. Dept., Brigham Young Univ., Provo, Utah, United States of America, 2002.
- [46] F. Block, A. Schmidt, N. Villar and H. Gellersen, "Towards a Playful User Interface for Home Entertainment Systems," in *2nd European Symp. Ambient Intelligence (EUSAI)*, Eindhoven, The Netherlands, 2004, pp. 207-217.
- [47] L. Nery e Silva, T. Aires and G. Lemos de Souza, "Desenvolvimento de programas de TVDI explorando as funções inovadoras do GINGA-J," in *14th Brazilian Symp. Multimedia and Web*, 2008, pp. 75-82.
- [48] S. Lenman, L. Bretzner and B. Eiderbäck, "Computer Vision Based Recognition of Hand Gestures for Human-Computer Interaction," Dept. Numerical Anal. Comput. Sci., Royal Inst. Technol., Stockholm, Sweden, Tech. Rep. CID-148, 2002.
- [49] J. E. P. Souza, D. M. Barreto and H. S. Pereira, "Acessibilidade para a TV Digital," in *Seminário de Informática - RS (SEMINFO RS'2008)*, Torres, Rio Grande do Sul, 2008.
- [50] IPTV Focus Group, "IPTV Focus Group Proceedings", International Telecommunications Union, Genoa, Suiza, 2008.
- [51] G. O'Driscoll, *Next Generation IPTV Services and Technologies*, New Jersey: John Wiley & Sons, Inc, 2008.
- [52] D. Ramirez, *IPTV Security: Protecting High Value Digital Contents*, Chichester, United Kingdom: John Wiley & Sons, 2008.
- [53] W. Simpson, *Video over IP: IPTV, Internet Video, H.264, P2P, Web TV, and Streaming: A Complete Guide to Understanding the Technology*, S. Memill, Ed., 2nd. ed., United States of America: Focal Press, 2008.
- [54] Grid Computing Focus Group, "Distributed Video-on-Demand – A grid based VoD Solution," Infosys Softw. Eng. and Technol. Labs., Bangalore, India, 2006.
- [55] B. Zhang, "A Large-Scale P2P Based Architecture for Video on Demand Service," M.S. thesis, Elect. and Comput. Eng. and Comput. Sci. Dept., Univ. Cincinnati, Cincinnati, OH, United States of America, 2007.
- [56] B. Añazco and L. Guzmán, "Análisis de arquitecturas existentes para sistemas VoD y diseño de Redes de acceso para proveer el servicio de video en demanda y de valor agregado para el Grupo TV Cable," thesis, Escuela Superior Politécnica del Litoral, Guayaquil, Ecuador, 2007.

- [57] T.D.C. Little and D. Venkatesh, "Prospects for Interactive Video-on-Demand," Multimedia Commun. Lab., Boston Univ., MA, United States of America, Tech. Rep. 02-15-1994, Aug., 2005.
- [58] F. Thouin and M. Coates, "Video-on-Demand Networks: Design Approaches and Future Challenges," *IEEE Network*, pp. 42-48, Marzo/Abril, 2007.
- [59] J. Balladini, "Un Sistema de Video bajo Demanda a Gran Escala Tolerante a Fallos de Red", Ph.D. tesis, Univ. Autónoma de Barcelona, Barcelona, España, 2008.
- [60] GoBackTV, Inc., "RetroVue™ System Overview," Menlo Park, CA, United States of America, 2007. Disponible: <http://www.gobacktv.com/pdf/overview-telco.pdf>
- [61] O. I. Silgado and J. L. Padilla, "Lineamientos Técnicos y Regulatorios para la Implementación de IPTV en Colombia," thesis, Univ. of Cauca, Popayán, Colombia, 2008.
- [62] C. E. Serrano, "Modelo para la Construcción de Soluciones," in *Modelo Integral para el Profesional en Ingeniería*, Ed. Universidad del Cauca, Popayán, Cauca, Colombia, 2005.
- [63] Asterisk – The Open Source Telephony Projects (2010, November 15th). [online] Available: <http://www.asterisk.org/home>
- [64] SpecsSheet – FreeSWITCH Wiki (2010, November 15th). [online] Available: <http://wiki.freeswitch.org/wiki/Specsheet>
- [65] Wiki – Callweaver (2010, November 15th). [Online] Available: <http://www.callweaver.org/wiki/CallWeaver>
- [66] Yate – the next generation telephony engine (2010, November 15th). [Online] Available: <http://yate.null.ro/pmwiki/index.php>
- [67] TransNexus, "Performance Benchmark Test for Asterisk B2BUA," TransNexus, Inc., Atlanta, GA, United States of America, 2008.
- [68] ITU-T Recommendation H.323, *Packet-based multimedia communications systems*, 2009.
- [69] L. Madsen, J. Meggelen and J. Smith, "Protocols for VoIP," in *Asterisk™: The Future of Telephony*, 2nd ed., M. Loukides, Ed. United States of America: O'Reilly Media, Inc., 2005, pp. 137-155.

- [70] H. Ingo, "Session Initiation Protocol (SIP) and other Voice over IP (VoIP) protocols and applications," Sesca Tech., Finland, 2007.
- [71] P. Nie, "An open standard for instant messaging: eXtensible Messaging and Presence Protocol (XMPP)," Helsinki Univ. Tech., Helsinki, Finland, May, 2006.
- [72] K. Mosquera and C. Muñoz, "Piloto de Sistema de Comunicación con VoWLAN Controlado por Comandos de Voz para un Entorno Hospitalario", thesis, Univ. Of Cauca, Popayán, Colombia, 2007.
- [73] Ixia, "Assessing VoIP Call Quality Using the E-model," West Agoura Road, Calabasas, CA, United States of America, 2005.
- [74] ITU-T Recommendation G.722, *General Aspects of Digital Transmission Systems*, 1998.
- [75] C. Collazos, J. Arciniegas, V. Mondragón and X. García, "Lineamientos de usabilidad para el diseño y evaluación de la televisión digital interactiva," *Avances Sist. Inform.*, vol. 5, no. 3, pp. 213-218, Dec., 2008.
- [76] J. Fernström, "Commercial Graphical User Interface for Digital TV," M.S. Thesis, Dept. Numerical Analysis and Comput. Sci., Royal Inst. Technol., Stockholm, Sweden, 2004.
- [77] C. Collazos, C. Rusu, J. Arciniegas, S. Roncagliolo, "Designing and Evaluating Interactive Television from a Usability Perspective," in *Proc 2009 2nd Int. Conf. Advances in Comput.-Human Interactions*, Cancun, Mexico, 2009, pp. 381-385.
- [78] N. Beese, "Designing for BBC Red Button—BBC Red Button & Interactive TV Programmes," British broadcasting Corporation, London, United Kingdom, 2009.
- [79] Telefónica I+D, "Taller imagenio," in *Campus Party 2008*, Valencia, España, 2008.
- [80] LumenVox, *LumenVox Speech Engine Support* (2010, November 15th). [online] Available: <http://www.lumenvox.com/help/speechEngine/administration/hardware-requirements.html>
- [81] *Measuring the Usability of Multi-Media Software* (2010, May 7th). [online]. Available: <http://www.ucc.ie/hfrg/questionnaires/mumms/info.html>
- [82] K. Hone and R. Graham, "Towards a Tool for the Subjective Assessment of Speech System Interfaces (SASSI)," *Natural Language Engineering*, vol. 6, pp. 287 – 303, Sept., 2000.

- [83] C. W. Turner, J. Nielsen and J. R. Lewis, "Current Issues in the Determination of Usability Test Sample Size: How Many Users is Enough?," in *Proceedings of Usability Professionals Association UPA '02*, Orlando, FL, United States of America, 2002.
- [84] J. Nielsen. (2010, June 15th). *Why You Only Need to Test with 5 Users* [online]. Available: <http://www.useit.com/alertbox/20000319.html>
- [85] ITU-T Recomendación Y.1541, *Objetivos de calidad de funcionamiento de red para servicios basados en el protocolo Internet*, 2006.
- [86] Cisco Systems, Inc., "Voice Over IP – Per Call Bandwidth Consumption," Cisco Systems, Inc., San Jose, CA, United States of America, Tech. Rep. 7934, 2006.
- [87] P. Quito, "Análisis de tráfico y costos de implementación de un sistema de VOIP mediante asterisk para Amadeus," Univ. Politécnica Salesiana, Cuenca, Ecuador, 2008.
- [88] ITU-T Recomendación G.114, *Tiempo de transmisión en un sentido*, 2003.
- [89] M. Darnell, "Acceptable system response times for TV and DVR," in *Proc. 5th European Conf. Interactive TV: a shared experience*, Amsterdam, The Netherlands, 2007, pp. 47-56.

Análisis del Uso de Interactividad basada en Voz para el manejo del Servicio de Video bajo Demanda

ANEXOS A, B y C

LAURA GÓMEZ FIGUEROA
JESÚS ARMANDO SANDOVAL LEON

Director
Ing. OSCAR JOSUÉ CALDERON CORTES

Trabajo de grado entregado como requisito para obtener el título de Ingeniero en
Electrónica y Telecomunicaciones

UNIVERSIDAD DEL CAUCA
FACULTAD DE INGENIERIA ELECTRONICA Y
TELECOMUNICACIONES
DEPARTAMENTO DE TELECOMUNICACIONES
GRUPO I+D EN NUEVAS TECNOLOGÍAS EN
TELECOMUNICACIONES

Análisis del Uso de Interactividad basada en Voz para el manejo del Servicio de Video bajo Demanda



ANEXO A

LAURA GÓMEZ FIGUEROA
JESÚS ARMANDO SANDOVAL LEON

UNIVERSIDAD DEL CAUCA
FACULTAD DE INGENIERIA ELECTRONICA Y TELECOMUNICACIONES
DEPARTAMENTO DE TELECOMUNICACIONES
GRUPO I+D EN NUEVAS TECNOLOGÍAS EN TELECOMUNICACIONES

ANEXO A. PROCESO DE INSTALACIÓN Y CONFIGURACIÓN

Dentro de este documento se describe el proceso de instalación y configuración de la solución de interactividad basada en VoIP para el manejo del servicio de VoD.

1. INSTALACIÓN DEL SISTEMA OPERATIVO

El sistema operativo elegido para realizar la instalación de los servidores es UBUNTU bajo la versión 10.04. Se elige este sistema operativo por ser software libre y no requerir licencia de instalación.

2. INSTALACION Y CONFIGURACION DE LOS SERVIDORES

En esta sección se explica el proceso de instalación del servidor de video y el servidor web, seguidos durante la primera actividad del proceso de configuración, implementación y puesta a punto de la solución.

2.1. INSTALACION Y CONFIGURACION DE DARWIN STREAMING SERVER (DSS)

2.1.1 Instalación

Para la correcta instalación de DSS se ejecuta el script "DSS-SCRIPT.sh", quien descarga e instala el servidor de video streaming automáticamente.

DSS-SCRIPT.sh

```
#!/bin/bash

#paquetes esenciales para la compilacion
sudo apt-get install build-essential wget

wget http://static.macosforge.org/dss/downloads/DarwinStreamingSrvr6.0.3-Source.tar
wget http://dss.macosforge.org/trac/raw-attachment/ticket/6/dss-6.0.3.patch
wget http://dss.macosforge.org/trac/raw-attachment/ticket/6/dss-hh-20080728-1.patch
wget http://dss.macosforge.org/trac/raw-attachment/ticket/6/Install

#descomprime los fuentes y crea el directorio DarwinStreamingSrvr6.0.3-Source
tar xf DarwinStreamingSrvr6.0.3-Source.tar
#aplica el patch: este patch es para hacerlo compatible con distribuciones de 64 bits
patch -p0 < dss-6.0.3.patch
#aplica el otro patch: este patch es para hacerlo compatible con distribuciones de 64 bits
patch -p0 < dss-hh-20080728-1.patch

#copia el nuevo archivo instalación al los fuentes reemplazando el existente
cp Install DarwinStreamingSrvr6.0.3-Source
#agrego el grupo y usuario qtss
sudo addgroup --system qtss
```

```
sudo adduser --system --no-create-home --ingroup qtss qtss
```

```
#entro a la carpeta
```

```
cd DarwinStreamingSrvr6.0.3-Source
```

```
#compilo
```

```
./Buildit
```

```
#instalo: aquí preguntara un nombre de usuario y contraseña de administración del servidor
```

```
sudo ./Install
```

Una vez se ha instalado de manera exitosa DSS, a través de consola se le indica al usuario que debe dirigirse a la dirección: <http://localhost:1220/> para configurar el servidor.

2.1.2 Configuración

Desde el navegador web, el administrador accede a la dirección <http://localhost:1220/>, la cual permite configurar los parámetros del servidor, entre las que se encuentran establecimiento de contraseñas, Figura A1, seguridad del servidor, Figura A2, puertos de transmisión, Figura A3, y la ubicación de las películas, Figura A4.



Figura A1. Establecimiento de la contraseña del servidor.



Figura A2. Configuración de la seguridad de acceso al servidor.

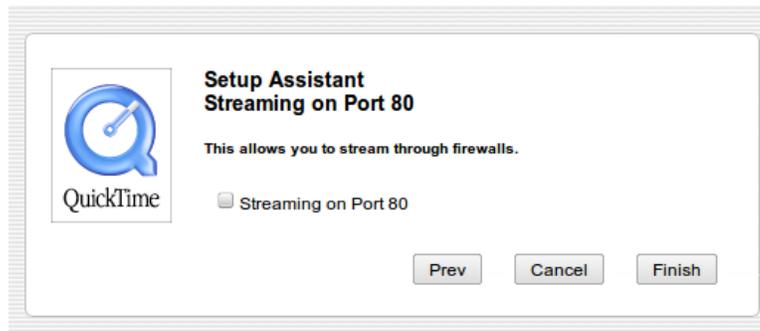


Figura A3. Configuración del puerto de transmisión.

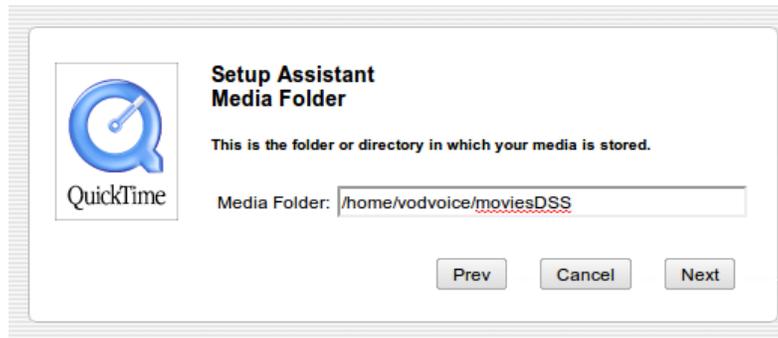


Figura A4. Configuración de la Ubicación de las películas.

Luego de la configuración del servidor, se accede por medio del a interfaz web para el monitoreo del servidor, para el que se requiere introducir el nombre de usuario y contraseña, establecidos durante el proceso de configuración, Figura A5.

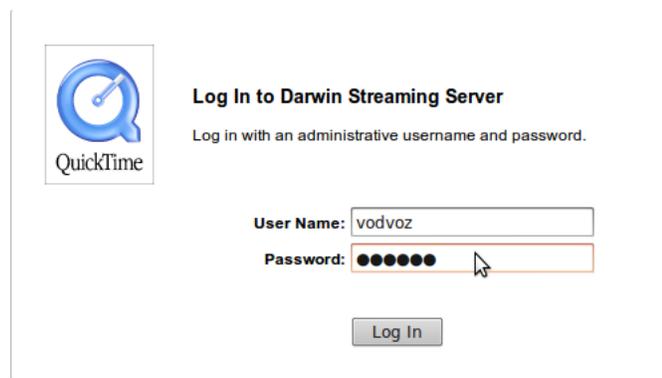


Figura A5. Autenticación para acceder al servidor DSS.

Luego del proceso de autenticación, se accede a la página web de monitoreo de DSS, Figura A6Figura A6, a través de la cual se pueden configurar los contenidos multimedia disponibles y se puede observar las sesiones establecidas, la tasa de transmisión, la pérdida de paquetes, etc.

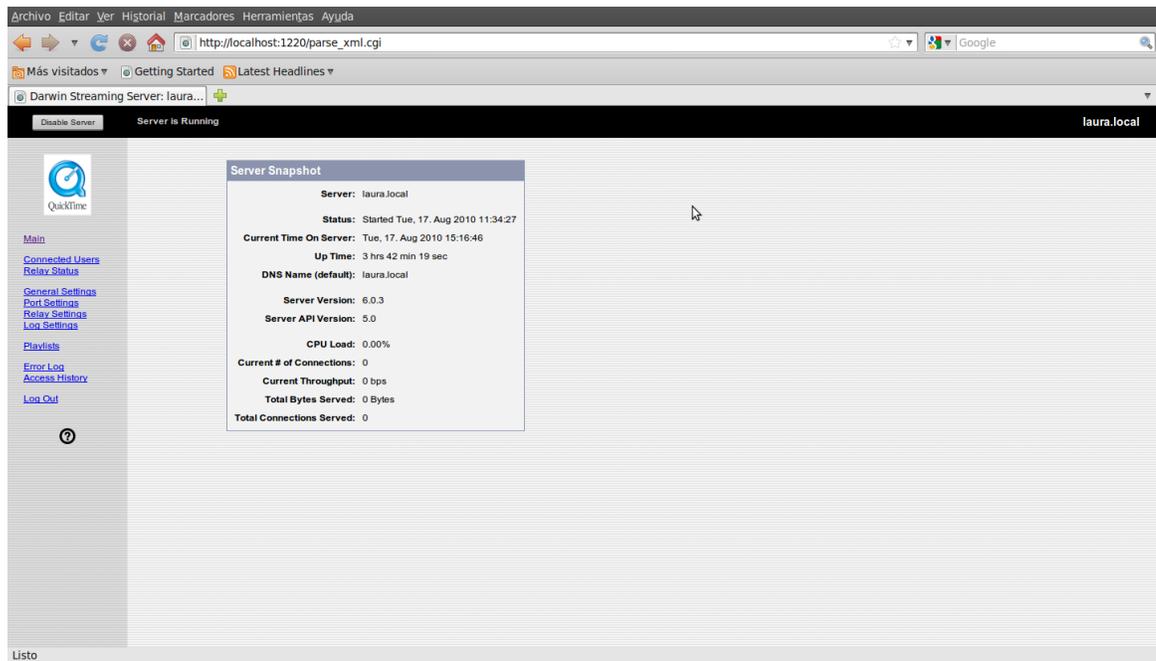


Figura A6. Interfaz Gráfica de monitoreo de DSS.

2.1.3 Preparación de los contenidos para Streaming en DSS

En esta sección se explica el proceso de preparación de los contenidos multimedia, proceso que se requiere realizar para que estos sean enviados a través de streaming.

El servidor Darwin Streaming Server soporta los códec de video H.261, H.263, **H.264**, MPEG-4 (MPEG-4 PARTE 2), entre otros. Entre los códec de audio soportados se encuentran AAC LC, AMR, etc. Se decide trabajar con el H.264, debido a su alta utilización en el mercado, además de la alta calidad de video con consumo de ancho de banda moderado. El códec de audio a utilizar es AAC.

La codificación de los contenidos multimedia se realiza por medio del SUPER¹⁷, Figura A7, herramienta multi-codec y muti-contenedor de sencilla utilización. Dicho programa usa el codec libre x264, para codificar en H.264/AVC.

¹⁷ Disponible en: <http://www.erightssoft.com/SUPER.html>

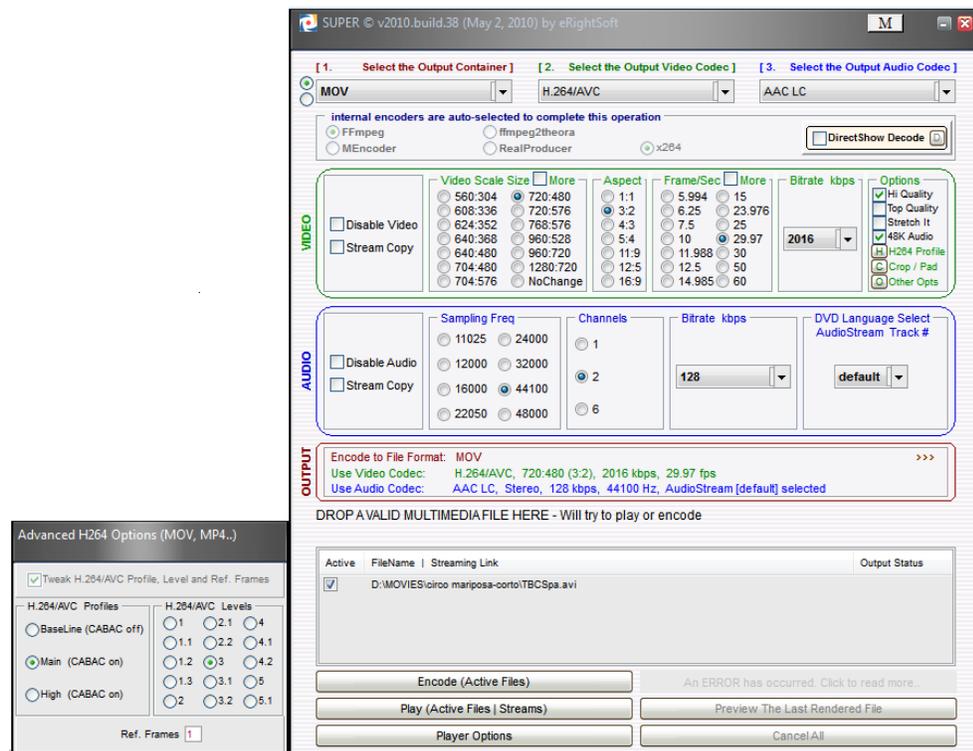


Figura A7. Programa de Codificación de los contenidos multimedia.

Luego de la codificación de los videos, se adicionan las pistas de indicación, necesarias para la transmisión de los contenidos multimedia en tiempo real, ya que le indican a DSS cómo empaquetar y enviar la información sobre la red de comunicaciones. Dichas indicaciones se agregan desde el Quick Time, a través de los siguientes pasos:

1. **Abrir la película:** Se abre la película codificada desde el Quick Time.
2. **Exportar la película:** Se selecciona **Archivo > Exportar > Película como Película con Indicaciones**, Figura A8, en este paso lo que se hace es, establecer indicaciones en el vídeo.
3. **Comprobar la adición de las pistas de indicaciones:** Se selecciona **Ventana > Mostrar propiedades de película**, Figura A9, en la que se observan además de las pistas de audio y video, las indicaciones de la pista de video y las indicaciones de la pista de audio.

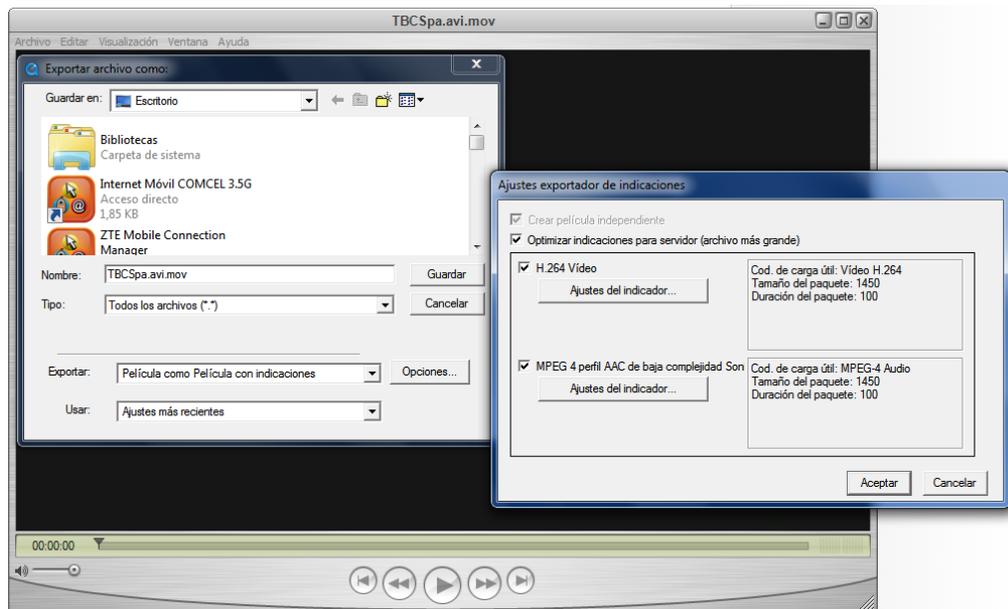


Figura A8. Indicaciones a la película a transmitir.

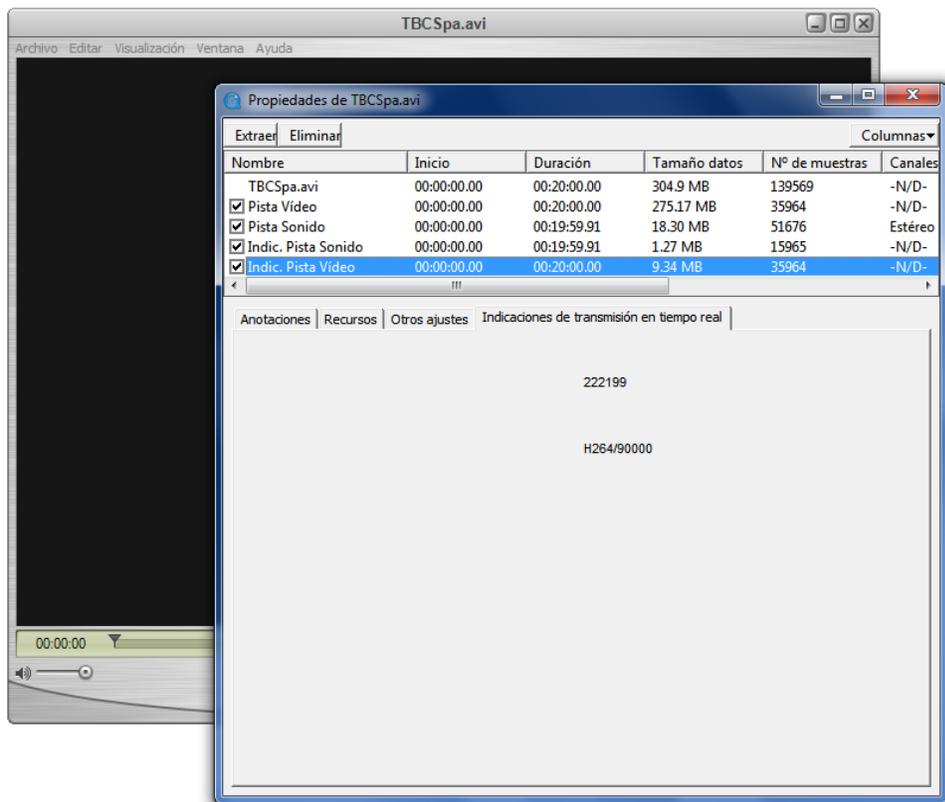


Figura A9. Contenido Multimedia con pistas de indicaciones.

3. INSTALACIÓN Y CONFIGURACIÓN DEL MÓDULO DE RECONOCIMIENTO DEL HABLA

En esta sección se explica la instalación y configuración de los elementos que integran el módulo de reconocimiento del habla, llevada a cabo durante la segunda actividad.

3.1 Instalación de ASTERISK

El primer elemento a instalar dentro del módulo de reconocimiento del habla, es el servidor de comunicaciones IP. A continuación, se describe el proceso de instalación del mismo.

3.1.1 Instalación de las dependencias

Previo a la instalación del servidor de comunicaciones IP, es necesario instalar las dependencias de Asterisk, ejecutando desde consola los siguientes comandos:

- ✓ *aptitude install gcc-4.4*
- ✓ *aptitude install dpatch build-essential*
- ✓ *aptitude install fakeroot patchutils*
- ✓ *aptitude install bison*
- ✓ *aptitude install libnewt0.52 libnewt-dev*
- ✓ *aptitude install openssl*
- ✓ *aptitude install libssl0.9.8 libssl-dev zlib1g-dev*
- ✓ *aptitude install libncurses5 libncurses5-dev*
- ✓ *aptitude install libncurses5 libncurses5-dev*

3.1.2 Descarga y Descompresión de los paquetes de ASTERIK

Una vez las dependencias se han instalado correctamente se procede a descargar las versiones estables de Asterisk y dahdi, disponible en: <http://www.asterisk.org/downloads> ó a través de consola ejecutando los siguientes comandos:

- ✓ *wget -c http://downloads.asterisk.org/pub/telephony/dahdi-linux-complete/releases/dahdi-linux-complete-2.2.1+2.2.1.tar.gz*
- ✓ *wget -c http://downloads.asterisk.org/pub/telephony/libpri/releases/libpri-1.4.10.2.tar.gz*
- ✓ *wget -c http://downloads.asterisk.org/pub/telephony/asterisk/releases/asterisk-1.6.0.25.tar.gz*
- ✓ *wget -c http://downloads.asterisk.org/pub/telephony/asterisk/releases/asterisk-addons-1.6.0.4.tar.gz*

Una vez los archivos han sido descargados, desde consola se abre la carpeta que los contiene y se ejecutan los siguientes comandos para descomprimirlos:

- ✓ `tar -xf dahdi-linux-complete-2.2.1+2.2.1.tar.gz`
- ✓ `tar -xf libpri-1.4.10.2.tar.gz`
- ✓ `tar -xf asterisk-1.6.0.25.tar.gz`
- ✓ `tar -xf asterisk-addons-1.6.0.4.tar.gz`

Luego de descomprimir los archivos, se procede a compilarlos e instalarlos.

3.1.3 Compilación e instalando DAHDI:

Para instalar el archivo dahdi, desde consola se ejecuta los siguientes comandos, quien abre la carpeta que ha sido descomprimida previamente:

`cd dahdi-linux-complete-2.2.1+2.2.1/`

Dentro de la carpeta, desde consola se ejecuta el comando make all y make install para compilar e instalar el archivo dahdi.

`make all`

`sudo make install`

Si este archivo ha sido instalado correctamente se debe obtener en la pantalla de la consola el siguiente mensaje:

```
#####  
###  
### DAHDI tools installed successfully.  
### If you have not done so before, install init scripts with:  
###  
### make config  
###  
#####
```

Finalmente, se realiza el proceso de configuración, ejecutando:

`sudo make config`

Para salir de carpeta de instalación actual, se ejecuta en consola "**`cd..`**".

3.1.4 Compilación e Instalación de LIBPRI:

Para instalar el archivo LIBPRI, se continúa un proceso similar al anterior, ejecutando los siguientes comandos:

```
cd libpri-1.4.10.2/  
make  
sudo make install  
cd ..
```

3.1.5 Compilación e Instalación de ASTERISK

Para instalar el servidor de comunicaciones IP, se instala ejecutando los siguientes comandos:

```
cd asterisk-1.6.0.25/
```

```
./configure
```

Si Asterisk se ha configurado correctamente, se debe obtener en consola:

```
.$$$$$$$$$$$$$$$$$$=..  
.$7$7..          .7$77:..  
          .$$:..          , $7.7  
.$7.          7$$$$          .$$77  
          ..$$          $$$$$          .$$$7  
          ..7$          .?.          $$$$$          .?.          7$$$.  
          $.$.          .$$$7.          $$$$7          .7$$$.  
          .777.          .$$$$$77$$$$77$$$$$7.          $$$,  
$$$~          .7$$$$$$$$$$$$$$$$7.          .$$$.  
.$$7          .7$$$$$$$$7:          ?$$$.  
$$$          ?7$$$$$$$$$$$I          .$$$7  
$$$          .7$$$$$$$$$$$$$$$$$          :$$$.  
$$$          $$$$$$7$$$$$$$$$$$$$          .$$$.  
$$$          $$$          7$$$7          .$$$          .$$$.  
$$$$          $$$7          .$$$.  
7$$$7          7$$$          7$$$  
$$$$$          $$$  
$$$$7.          $$ (TM)  
$$$$$.          .7$$$$$ $$  
$$$$$$$$$7$$$$$$$$$.$$$$$  
$$$$$$$$$$$$$.
```

```
configure: Package configured for:  
configure: OS type : linux-gnu  
configure: Host CPU : i686  
configure: build-cpu:vendor:os: i686 : pc : linux-gnu :  
configure: host-cpu:vendor:os: i686 : pc : linux-gnu :
```

```
configure: Package configured for:
```



```
Installing other config files...
Installing file configs/adsi.conf.sample
.....
.....
Installing file phoneprov/000000000000-phone.cfg
Installing file phoneprov/polycom.xml
```

sudo make config

Si este comando se ejecuta correctamente se debe obtener:

```
update-rc.d: warning: /etc/init.d/asterisk missing LSB information
update-rc.d: see <http://wiki.debian.org/LSBInitScripts>
Adding system startup for /etc/init.d/asterisk ...
/etc/rc0.d/K91asterisk -> ../init.d/asterisk
/etc/rc1.d/K91asterisk -> ../init.d/asterisk
/etc/rc6.d/K91asterisk -> ../init.d/asterisk
/etc/rc2.d/S50asterisk -> ../init.d/asterisk
/etc/rc3.d/S50asterisk -> ../init.d/asterisk
/etc/rc4.d/S50asterisk -> ../init.d/asterisk
/etc/rc5.d/S50asterisk -> ../init.d/asterisk
```

cd ..

3.1.6 Instalación y Configuración de ASTERISK ADDONS

Para instalar y configurar este archivo, se ejecutan los siguientes comandos:

cd asterisk-addons-1.6.0.4/

./configure :

Si este comando se ejecuta correctamente se debe obtener:

```

      .$$$$$$$$$$$$$$$$$=..
      .7$77..           .7$$77:.
      .$$:..           , $7.7
.$7.      7$$$$$      .$$77
      ..$$$.      $$$$$$      .$$$7
      ..7$  .?.  $$$$$$  .?.  7$$$$.
      $.$.  .$$$7.  $$$$7  .7$$$$.  .$$$$.
      .777.  .$$$$$$77$$$$77$$$$$7.  $$$$,
      $$$~  .7$$$$$$$$$$$$$$7.  .$$$$.
      .$$7  .7$$$$$$$$$7:  ?$$$$.
      $$$  ?7$$$$$$$$$$$$I  .$$$7
      $$$  .7$$$$$$$$$$$$$$$$$  :$$$$.
      $$$  $$$$$$7$$$$$$$$$$$$$  .$$$$.
      $$$  $$$  7$$$$7  .$$$  .$$$$.
      $$$$  $$$$7  .$$$7  .$$$$.
```

```
7$$$7          7$$$$          7$$$
$$$$$          $$$
$$$$7.         $$ (TM)
$$$$$$$.      .7$$$$$$ $$
$$$$$$$$$$$$7$$$$$$$$$.$$$$$$
$$$$$$$$$$$$$$$$$.
```

```
configure: Package configured for:
configure: OS type : Linux
configure: Host CPU : i686
```

make

Si este comando se ejecuta correctamente se debe obtener:

```
+---- Asterisk-Addons Build Complete ----+
+
+ Addons has successfully been built .      +
+ If you would like to install it :        +
+                                           +
+                make install              +
+-----+
```

sudo make install

Si este comando se ejecuta correctamente se debe obtener:

```
+---- Asterisk-Addons Installation Complete ----+
+
+ Addons has successfully been installed.    +
+ If you would like to install the sample   +
+ configuration files (overwriting any     +
+ existing config files), run:            +
+                                           +
+                make samples              +
+-----+
```

make samples

Con el anterior comando, concluye el proceso de instalación del servidor de comunicaciones IP, queda correctamente instalado.

3.1.7 Archivos de Configuración

Para la configuración del servidor de comunicaciones se deben modificar los archivos sip.conf, extensions.conf y modules.conf. El archivo modules.conf se modifica una vez el motor de reconocimiento se ha instalado. Los archivos empleados en la solución, se detallan a continuación:

SIP.CONF

/etc/asterisk/sip.conf

```
[general]
context=default
allowoverlap=no
bindport=5060
bindaddr=0.0.0.0
tcpenable=no
tcpbindaddr=0.0.0.0
srvlookup=yes
disallow=all
allow=ulaw
allow=alaw
allow=ilbc
allow=gsm
allow=h261

language=es
; Definicion de los usuarios

[armando]
type=friend
secret=armando
callerid=armando
qualify=yes
nat=no
host=dynamic
canreinvite=no
context=internal

[perseo]
type=friend
secret=perseo
callerid=perseo
qualify=yes
nat=no
host=dynamic
canreinvite=no
context=internal
port=5061 ;

[zeus]
type=friend
secret=zeus
callerid=zeus
qualify=yes
nat=no
host=dynamic
```

```
canreinvite=no
context=internal
```

EXTENSIONS.CONF

/etc/asterisk/extensions.conf

```
[general]
static=yes
writeprotect=no
clearglobalvars=no
```

```
[internal]
```

```
;ExTENSION DE PRUEBA
```

```
exten => 201,1,answer()
exten => 201,2,sayDigits("1234")
exten => 201,3,playback(demo-echotest)
exten => 201,4,echo
exten => 201,5,playback(demo-echodone)
exten => 201,6,wait(5)
exten => 201,7,hangup()
```

```
;*****
*****
exten => 6,1,Answer
exten =>
6,n,Set(UMBRAL=750);*****
*****
exten => 6,n,SpeechCreate
exten =>
6,n(cargaI),SpeechLoadGrammar(inicial,/etc/asterisk/gramatica/inicial.grxml)
;*****
*****
exten => 6,n,SpeechActivateGrammar(inicial)
exten => 6,n(beep2),SpeechBackground(beep)
exten => 6,n,Verbose(1,reconocido fue
${SPEECH_TEXT(0)});*****
exten => 6,n,Verbose(1,con una confianza de
${SPEECH_SCORE(0)});*****
exten => 6,n,GotoIf("${SPEECH_SCORE(0)}" >
"${UMBRAL}")?gif2:beep2);*****
exten => 6,n(gif2),GotoIf("${SPEECH_TEXT(0)}" =
"empezarserv"?agi2:beep2)
exten => 6,n(agi2),Agi(agi://localhost/socket2.agi)
exten => 6,n,SpeechDeactivateGrammar(inicial)
exten => 6,n,SpeechUnloadGrammar(inicial);

exten =>
6,n(menu),SpeechLoadGrammar(menu,/etc/asterisk/gramatica/menu.grxml);
exten => 6,n,SpeechActivateGrammar(menu)
```

Análisis del Uso de Interactividad basada en Voz para el manejo del Servicio de Video bajo Demanda

```
exten => 6,n(beep3),SpeechBackground(beep)
exten => 6,n,Verbose(1,reconocido fue
${SPEECH_TEXT(0)});*****
exten => 6,n,Verbose(1,con una confianza de
${SPEECH_SCORE(0)});*****
exten => 6,n,GotoIf("${SPEECH_SCORE(0)}" >
"${UMBRAL}")?agium1:beep3);*****

exten => 6,n(agium1),Agi(agi://localhost/socket2.agi)
exten => 6,n,GotoIf("${SPEECH_TEXT(0)}" = "ver"?desm2:if2); si es
verdadero va a desm2 sino va a if2
exten => 6,n(if2),GotoIf("${SPEECH_TEXT(0)}" =
"desactivado"?desm1:beep3)
exten => 6,n(desm1),SpeechDeactivateGrammar(menu)
exten => 6,n,SpeechUnloadGrammar(menu);

;*****
exten =>
6,n,SpeechLoadGrammar(activar,/etc/asterisk/gramatica/conactivacion/activ
ar.grxml)
exten => 6,n,SpeechActivateGrammar(activar)
exten => 6,n(beep4),SpeechBackground(beep)

exten => 6,n,Verbose(1,reconocido fue
${SPEECH_TEXT(0)});*****
exten => 6,n,Verbose(1,con una confianza de
${SPEECH_SCORE(0)});*****
exten => 6,n,GotoIf("${SPEECH_SCORE(0)}" >
"${UMBRAL}")?gif3:beep4);*****

exten => 6,n(gif3),GotoIf("${SPEECH_TEXT(0)}" = "activado"?agi3:beep4)
exten => 6,n(agi3),Agi(agi://localhost/socket2.agi)
exten => 6,n,SpeechDeactivateGrammar(activar)
exten => 6,n,SpeechUnloadGrammar(activar);
;*****
***
exten => 6,n,Goto(menu)

exten => 6,n(desm2),SpeechDeactivateGrammar(menu)
exten => 6,n,SpeechUnloadGrammar(menu);

exten =>
6,n(repro),SpeechLoadGrammar(reproduccion,/etc/asterisk/gramatica/reprodu
ccion.grxml)
exten => 6,n,SpeechActivateGrammar(reproduccion)
exten => 6,n(beep5),SpeechBackground(beep)
exten => 6,n,NoOp(LA HORA beep5 ES: ${STRFTIME(${EPOCH},,%Y%m%d-%H-%M-
%S)});*****

exten => 6,n,Verbose(1,reconocido fue
${SPEECH_TEXT(0)});*****
exten => 6,n,Verbose(1,con una confianza de
${SPEECH_SCORE(0)});*****
exten => 6,n,GotoIf("${SPEECH_SCORE(0)}" >
"${UMBRAL}")?agium2:beep5);*****
```

```
exten => 6,n(agium2),Agi(agi://localhost/socket2.agi)
exten => 6,n,NoOp(LA HORA agium2 ES: ${STRFTIME(${EPOCH},,%Y%m%d-%H-%M-%S)});*****
exten => 6,n,GotoIf("${SPEECH_TEXT(0)}" = "vinicial"?desr1:if3)
exten => 6,n(if3),GotoIf("${SPEECH_TEXT(0)}" = "desactivado"?desr2:beep5)
exten => 6,n(desr2),SpeechDeactivateGrammar(reproduccion)
exten => 6,n,SpeechUnloadGrammar(reproduccion);
;*****
exten =>
6,n,SpeechLoadGrammar(activar,/etc/asterisk/gramatica/conactivacion/activar.grxml)
exten => 6,n,SpeechActivateGrammar(activar)
exten => 6,n(beep6),SpeechBackground(beep)
exten => 6,n,NoOp(LA HORA beep6 ES: ${STRFTIME(${EPOCH},,%Y%m%d-%H-%M-%S)});*****

exten => 6,n,Verbose(1,reconocido fue ${SPEECH_TEXT(0)});*****
exten => 6,n,Verbose(1,con una confianza de ${SPEECH_SCORE(0)});*****
exten => 6,n,GotoIf("${SPEECH_SCORE(0)}" > "${UMBRAL}")?gif4:beep6);*****

exten => 6,n(gif4),GotoIf("${SPEECH_TEXT(0)}" = "activado"?agi4:beep6)
exten => 6,n(agi4),Agi(agi://localhost/socket2.agi)
exten => 6,n,NoOp(LA HORA agi4 ES: ${STRFTIME(${EPOCH},,%Y%m%d-%H-%M-%S)});*****
exten => 6,n,SpeechDeactivateGrammar(activar)
exten => 6,n,SpeechUnloadGrammar(activar);
;*****
***
exten => 6,n,Goto(repro)
exten => 6,n(desr1),SpeechDeactivateGrammar(reproduccion)
exten => 6,n,SpeechUnloadGrammar(reproduccion);
exten => 6,n,Goto(menu)
```

MODULES.CONF

/etc/asterisk/modules.conf

```
[modules]
autoload=yes
preload => func_strings.so
noload => pbx_gtkconsole.so
load => res_musiconhold.so
noload => chan_alsa.so
noload => chan_console.so

;carga de los modulos de reconocimiento
load => res_speech.so
load => res_speech_lumenvox.so
```

3.2 Instalación de LUMENVOX

En esta sección se explica el proceso de instalación del motor de reconocimiento, quien es el siguiente elemento dentro del módulo de reconocimiento del habla.

3.2.1 Instalación de las Dependencias

Para la instalación de Lumenvox (versión 9.0.601) en Ubuntu 10.04, se deben instalar las siguientes dependencias:

- ✓ libboost-filesystem1.34.1_1.34.1-14_i386
- ✓ libboost-regex1.34.1_1.34.1-14_i386
- ✓ libboost-thread1.34.1_1.34.1-14_i386
- ✓ libicu38_3.8-6ubuntu0.2_i386
- ✓ libmozjs1d_1.9.0.18-1_i386

3.2.2 Instalación

Para instalar el motor de reconocimiento se deben descargar los siguientes paquetes de instalación, disponibles en: <http://www.lumenvox.com/packages/Debian/binary/>

- ✓ lumenvoxcore_9.0.601_i386
- ✓ lumenvoxclient_9.0.601_i386
- ✓ lumenvoxlicenseserver_9.0.601_i386
- ✓ lumenvoxsre_9.0.601_i386

Todos estos paquetes se instalan por medio de la herramienta gdebi incluida en Ubuntu.

3.2.3 Instalación de la licencia

La licencia se adquiere desde la página de LumenVox, y ésta se encuentra asociada a la MAC del equipo servidor en el cual está instalado el motor.

Para instalar la licencia hay que crear un archivo de identificación del servidor, el cual llevara la información del equipo donde se instalará el motor de reconocimiento. Éste archivo se genera de la siguiente manera:

1. Asegurarse que el lv_license_server está corriendo. Si no es así, se ejecuta: **sudo /etc/init.d/lvlicensed start**
2. Crear el archivo ID del servidor: para ello se ejecuta: **/usr/bin/lv_license_manager -g path** siendo path el lugar donde se generará el archivo ID del servidor, por ejemplo: **/usr/bin/lv_license_manager -g /home/vodvoice/info.bts**

Si este proceso se realiza correctamente, se debe obtener: **Server hardware info saved to/home/vodvoice/info.bts successfully**

3. Descargar la licencia: para ello se dirige a la página web de Lumenvox, y se inicia sesión con los valores asignados durante la compra de la licencia, Figura A10, en la cual se presiona “ver licencia” para acceder a la página que permite descargarla.



Figura A10. Página web de LumenVox.

La Figura A11, muestra la página a la cual se debe acceder para solicitar la licencia. Dicha página, solicita subir el archivo identificador del servidor que ha sido generado previamente.

Una vez se sube a la página el archivo **info.bts**, es posible acceder la licencia del motor de reconocimiento, Figura A12, la cual se descarga luego de aceptar los términos establecidos en la licencia de uso del motor, Figura A13.

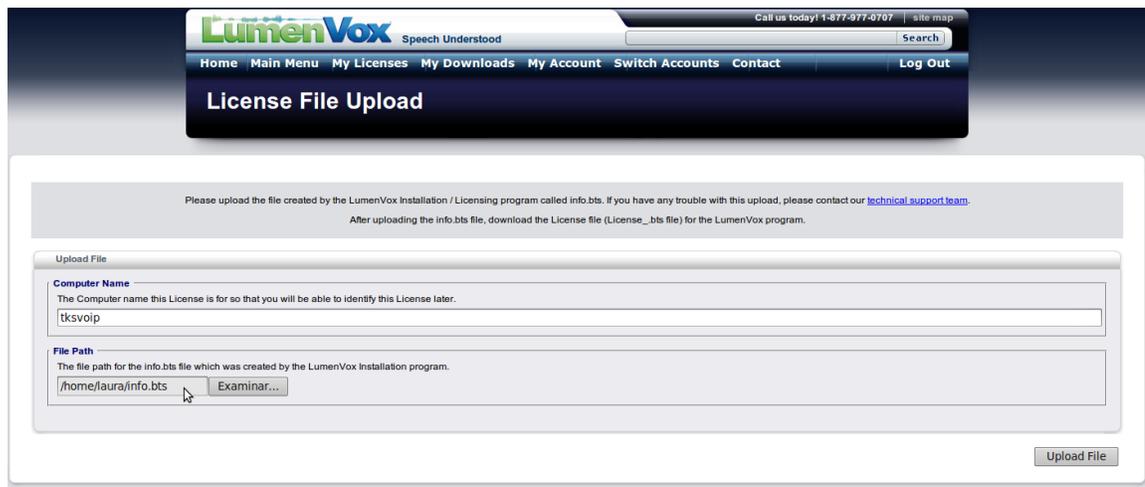


Figura A11. Solicitud de la Licencia.



Figura A12. Descargar Licencia.

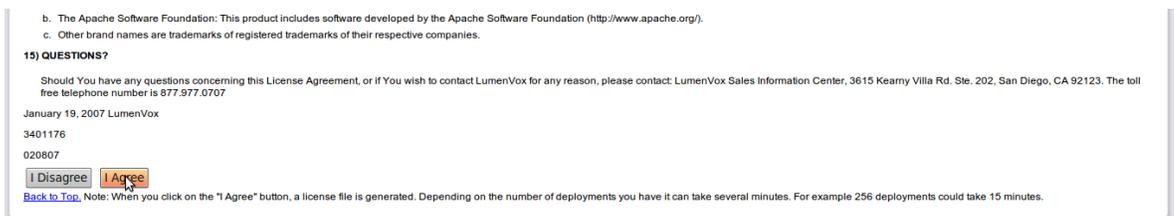


Figura A13. Licencia de Uso del Motor de Reconocimiento.

Luego de realizar el proceso de descarga de la licencia, se obtiene el archivo: License1812_r2.bts.

- 4. Instalación de la licencia:** para instalar la licencia del motor de reconocimiento se ejecuta en consola: `/usr/bin/lv_license_manager -m Licensefile` donde Licensefile es la ruta donde se encuentra almacenado el archivo de licencia descargado, por ejemplo:

`/usr/bin/lv_license_manager -m /home/vodvoice/License1812_r2.bts`

Si la licencia se ha instalado correctamente se debe obtener:
Install license [/home/vodvoice/License1812_r2.bts] successfully

3.2.4 Instalar el conector ASTERISK – LUMENVOX

Para integrar Asterisk y LumenVox se debe instalar el conector: connector-Bridge, el cual se encuentra disponible en:

<https://lumenvox.com/Downloads/asterisk-1.6.0-lumenvox-support-linux-i686-32bit-b24-engine9.0.tar.gz>

Éste archivo se extrae y se copia el archivo `res_speech_lumenvox.so` a la carpeta `modules` de Asterisk, lo cual se realiza ejecutando en consola:

```
mv./res_speech_lumenvox.so/usr/lib/asterisk/modules/res_speech_lumenvox.so
```

Así mismo, se debe mover el archivo `lumenvox.conf` a `/etc/Asterisk`, ejecutando en consola:

```
mv ./lumenvox.conf /etc/asterisk/lumenvox.conf
```

Luego de copiar el conector, se deben editar el archivo `/etc/lumenvox/client_property.conf` y establecer la variable en 1: `MODULE_STARTUP_MODE = 1`.

CLIENT_PROPERTY.CONF

```
/etc/lumenvox/client_property.conf
```

```
SRE_SERVERS = 127.0.0.1  
LICENSE_TYPE = Auto  
LIC_SERVER_HOSTNAME = 127.0.0.1  
LIC_SERVER_PORTNUM = 7569  
MAX_NBEST_RETURNED = 1  
DECODE_THREAD_PRIORITY = 1  
DECODE_TIMEOUT = 20000  
LOAD_GRAMMAR_TIMEOUT = 200000  
STRICT_SISR_COMPLIANCE = 0  
TRIM_SILENCE_VALUE = 970  
SAVE_SOUND_FILES = 0  
LANGUAGE = es-CO  
SEARCH_BEAM_WIDTH = 1e-6  
DECODE_OPTIMIZATION = 0  
NOISE_REDUCTION_ENABLE = 1  
MODULE_STARTUP_MODE = 1
```

SRE_SERVER.CONF

```
/etc/lumenvox/sre_server.conf
```

```
HMM_TYPE = SEMI
```

```
ENABLE_CACHE = 1
CACHE_EXPIRATION = 1440
CACHE_MAX_NUMBER = 100
CACHE_MAX_MEMORY = 268435456
SRE_PORTNUM = 5730
NUM_DECODE_THREADS = 4
```

Luego de este proceso, Asterisk y LumenVox se encuentran conectados. Por tanto, la siguiente etapa es la adición de las gramáticas del sistema, a las cuales accederá el motor de reconocimiento.

3.2.4 Gramáticas del Sistema

Dentro de la solución, se utilizan 4 gramáticas: Inicial.grxml, Menu.grxml, Reproduccion.grxml y Activar.grxml. Estas gramáticas se construyen con entrevistas informales en las cuales se indaga acerca de los botones del control remote más utilizados durante la interacción con el DVD.

INICIAL.GRXML

```
<?xml version="1.0"?>
<grammar xmlns="http://www.w3.org/2001/06/grammar"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.w3.org/2001/06/grammar
    http://www.w3.org/TR/speech-grammar/grammar.xsd"
  xml:lang="es-CO" version="1.0"
  root="Iniciales"
  mode="voice"
  tag-format="semantics/1.0.2006">
<rule id="Inicial">
  <one-of>
    <item>
      <one-of>
        <item>iniciar servicio</item>
        <item>empezar</item>
      </one-of>
      <tag>out="empezarserv"</tag>
    </item>
  </one-of>
</rule>
<rule id="Iniciales">
  <item>
    <tag>out=";"</tag>
  </item>
  <item repeat="1-">
    <ruleref uri="#Inicial"/>
    <tag>out=rules.latest();</tag>
  </item>
</rule>
</grammar>
```

MENU.GRXML

```
<?xml version="1.0"?>
<grammar xmlns="http://www.w3.org/2001/06/grammar"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.w3.org/2001/06/grammar
    http://www.w3.org/TR/speech-grammar/grammar.xsd"
  xml:lang="es-CO" version="1.0"
  root="Menus"
  mode="voice"
  tag-format="semantics/1.0.2006">
<rule id="Menu">
  <one-of>
<!-- comando para descativar reconocimiento(desactiva esta gramatica) -->
    <item>
      <one-of>
        <item>desactivar reconocimiento de voz</item>
        <item>desactivar comandos de voz</item>
      </one-of>
      <tag>out="desactivado"</tag>
    </item>
<!-- mostrar la siguiente pelicula -->
    <item>
      <one-of>
        <item>adelante</item>
<!--      <item>derecha</item> -->
<!--      <item>siguiente</item> -->
      </one-of>
      <tag>out="adelante"</tag>
    </item>
<!-- mostrar la anterior pelicula -->
    <item>
      <one-of>
        <item>atras</item>
<!--      <item>izquierda</item> -->
<!--      <item>anterior</item> -->
      </one-of>
      <tag>out="atras"</tag>
    </item>
<!-- iniciar la reproduccion de la pelicula -->
    <item>
      <one-of>
        <item>ver pelicula</item>
        <item>pelicula</item>
      </one-of>
      <tag>out="ver"</tag>
    </item>
<!-- volver al menu anterior -->
    <item>
      <one-of>
        <item>ver animacion</item>
        <item>animacion</item>
      </one-of>
      <tag>out="animacion"</tag>
    </item>
  </one-of>
</rule>
```

```
<item>
  <one-of>
    <item>ver cortometrajes</item>
    <item>cortometrajes</item>
  </one-of>
  <tag>out="cortometraje"</tag>
</item>
  <item>
    <one-of>
      <item>ayuda</item>
    </one-of>
    <tag>out="ayuda"</tag>
  </item>
</one-of>
</rule>
<rule id="Menus">
  <item>
    <tag>out="";</tag>
  </item>
  <item repeat="1-">
    <ruleref uri="#Menu"/>
    <tag>out=rules.latest();</tag>
  </item>
</rule>
</grammar>
```

REPRODUCCION.GRXML

```
<?xml version="1.0"?>
<grammar xmlns="http://www.w3.org/2001/06/grammar"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.w3.org/2001/06/grammar
    http://www.w3.org/TR/speech-grammar/grammar.xsd"
  xml:lang="es-CO" version="1.0"
  root="Repros"
  mode="voice"
  tag-format="semantics/1.0.2006">
<rule id="Repro">
  <one-of>
<!-- comando para descativar reconocimiento(desactiva esta gramatica) -->
    <item>
      <one-of>
        <item>desactivar reconocimiento de voz</item>
        <item>desactivar comandos de voz</item>
      </one-of>
      <tag>out="desactivado"</tag>
    </item>
    <item>pausar<tag>out="pausar"</tag></item>
<!-- volver al menu anterior -->
    <item>
      <one-of>
        <item>menu principal</item>
      </one-of>
      <tag>out="vinicial"</tag>
    </item>
```

```
<item>
  <one-of>
    <item>plai</item>
  </one-of>
  <tag>out="play"</tag>
</item>
<item>
  <one-of>
    <item>aumentar volumen</item>
    <item>subir volumen</item>
    <item>mas volumen</item>
  </one-of>
  <tag>out="masvol"</tag>
</item>
<item>
  <one-of>
    <item>disminuir volumen</item>
    <item>bajar volumen</item>
    <item>menos volumen</item>
  </one-of>
  <tag>out="menvol"</tag>
</item>
  <item>
    <one-of>
      <item>silencio</item>
    </one-of>
    <tag>out="silencio"</tag>
  </item>
</one-of>
</rule>
<rule id="Repros">
  <item>
    <tag>out="";</tag>
  </item>
  <item repeat="1-">
    <ruleref uri="#Repro"/>
    <tag>out=rules.latest();</tag>
  </item>
</rule>
</grammar>
```

ACTIVAR.GRXML

```
<?xml version="1.0"?>
<grammar xmlns="http://www.w3.org/2001/06/grammar"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.w3.org/2001/06/grammar
    http://www.w3.org/TR/speech-grammar/grammar.xsd"
  xml:lang="es-CO" version="1.0"
  root="Activar"
  mode="voice"
  tag-format="semantics/1.0.2006">
<rule id="Inicial">
  <one-of>
```

```
<item>
  <one-of>
    <item>activar reconocimiento de voz</item>
    <item>activar comandos de voz</item>
  </one-of>
  <tag>out="activado"</tag>
</item>
</one-of>
</rule>
<rule id="Activar">
  <item>
    <tag>out=";"</tag>
  </item>
  <item repeat="1">
    <ruleref uri="#Inicial"/>
    <tag>out=rules.latest();</tag>
  </item>
</rule>
</grammar>
```

Con la configuración de las gramáticas del sistema, se concluye configuración del motor de reconocimiento.

3.2.5 Funcionamiento Interto del Dial Plan

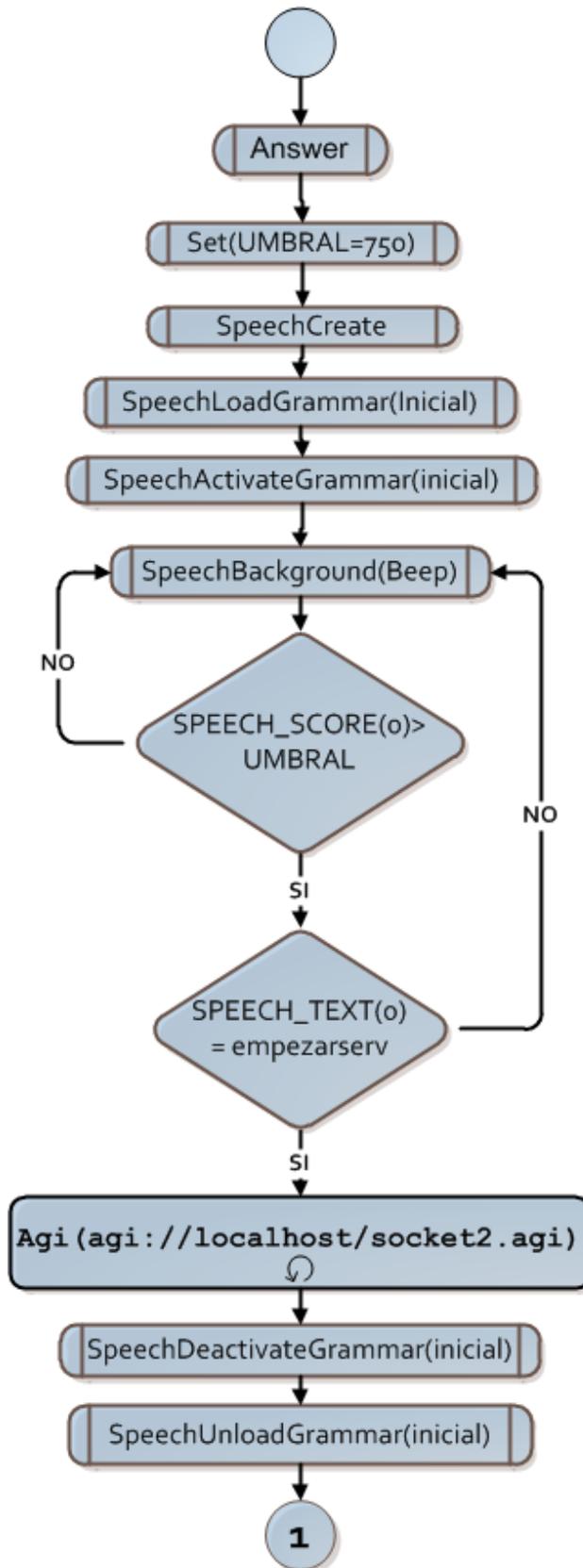
La figura A14, muestra la logica del dial plan del sistema en servidor de comunicaciones. Aquí se establece el umbral de reconocimiento. Un comando reconocido tiene a su vez una calificación de reconocimiento, entonces una vez se reconoce un comando se compara dicha calificación con el umbral, si el comando reconocido no pasa el umbral, entonces no se envía, esto en búsqueda de mejorar la precisión y disminuir la probabilidad de comandos erróneos. Además se dan los procesos de activación y desactivación de las gramáticas cuando se pasa una interfaz a otra.

Al tener gramáticas pequeñas, se disminuye la posibilidad de que se reconozca mal un comando, esto en comparación con el caso de una sola gramática para todo el sistema.

El mecanismo de activación y desactivación del motor de reconocimiento (se realiza una vez el usuario dice el comando "desactivar reconocimiento de voz", entonces se desactiva la gramática de la interfaz en la que se encuentra y se activa la gramática con un único comando: "activar reconocimiento de voz".

El envío de un comando se realiza por medio del AGI, y es aquí, en el dialplan, donde se hace el llamado a la AGI una vez un comando ha sido reconocido con éxito. El AGI simplemente captura el comando reconocido y lo envía al applet.

La figura A14, muestra un ciclo infinito, el cual inicia cuando se contesta la llamada en el servidor de comunicaciones. Si se desea terminar el ciclo, se debe colgar la llamada desde el softphone o el teléfono IP, de lo contrario el motor de reconocimiento quedará a la espera de un nuevo comando.



La AGI se ejecuta y pone en funcionamiento a través del comando: **java -jar asterik-java.jar**

AGIFINAL2.JAVA

```
import org.asteriskjava.fastagi.AgiChannel;
import org.asteriskjava.fastagi.AgiException;
import org.asteriskjava.fastagi.AgiRequest;
import org.asteriskjava.fastagi.BaseAgiScript;
import java.io.*;
import java.net.*;
publicclass AgiFinal2 extends BaseAgiScript {
publicvoid service(AgiRequest request, AgiChannel channel)
throws AgiException {
    String text = null;
    ServerSocket listen_socket;
text = this.getFullVariable("${SPEECH_TEXT(0)}");
System.out.println("lo reconocido fue " + text);
Socket misocket = null;
try {
misocket = new Socket("10.1.1.4", 6700);
System.out.println("se creo el socket ");
PrintStream soutstream = newPrintStream(misocket.getOutputStream
System.out.println("se creo el stream de salida");
soutstream.println(text);
System.out.println("Se envio " + text);
} catch (IOException e) {
System.err.println(e);
} finally {
try {
if (misocket != null) {
misocket.close();
} catch (IOException e) {
}
}
}
```

El archivo **fastagi-mapping.properties** enlaza el archivo asterik-java con AgiFinal2.java

FASTAGI-MAPPING.PROPERTIES

```
socket2.agi = AgiFinal2
```

Con la configuración de la AGI, se concluye la segunda actividad del proceso de instalación y configuración de la solución propuesta.

4. CLIENTE

En esta sección, se explica la instalación y configuración de los elementos del módulo cliente, correspondientes a la tercera y última actividad de la fase de configuración, implementación y puesta a punto.

4.1 Interfaz Gráfica de Usuario

Para el desarrollo de la interfaz gráfica de usuario se tuvieron en cuenta algunos principios de diseño, como tamaño de la letra, color de fondo, etc. La funcionalidad se implementa con HTML, Flash y JavaScript. La decodificación y control del video Streaming se realiza mediante el complemento de VLC para navegadores WEB. La Figura A15, muestra la distribución de la Interfaz de usuario.

La primera página (**index.html**) que aparece al iniciar el servicio es la de introducción, en ella se le da información al usuario acerca del uso del servicio. La estructura de dicha página es:

Marco superior: donde se carga el Applet (**Applet.html**), quien es la interfaz entre la GUI y el módulo de reconocimiento del habla.

Marco inferior: se muestran las páginas web utilizadas para la guía de contenidos y el reproductor multimedia. Ésta se construye por medio de un slider de imágenes 3D en flash **cu3er.swf**, disponible en: <http://getcu3er.com/>



Figura A15. Interfaz Gráfica de Usuario.

A continuación, se especifican los archivos utilizados para la GUI de la solución al iniciar el servicio.

INDEX.HTML

```
<!doctype html public "-//w3c//dtd html 4.0 transitional//en">
```

```
<html>
  <head>
    <meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
    <meta name="GENERATOR" content="Mozilla/4.5 [en] (WinNT; I) [Netscape]">
    <title>:-: VoDiVoice :-:</title>
  </head>
  <frameset ROWS="8%,*" COLS="25%" border="0" frameborder="0">
    <frame NAME="applet" SRC="http://192.168.120.156/~laura/applet.html" scrolling="no"
noreferrer="noreferrer">
    <frame NAME="intro" SRC="http://192.168.120.156/~laura/intro1.html" scrolling="no"
noreferrer="noreferrer">
  </frameset>
</html>
```

APPLET.HTML

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1" />
    <title>Applet de prueba</title>
    <script type="text/javascript" src="voicevod.js">
    </script>
    <script type="text/javascript" src="javascript/FlashHandler.js"></script>
  </head>
  <body bgcolor=#000000' TOPMARGIN='0' LEFTMARGIN='0' MARGINWIDTH='0'
MARGINHEIGHT='0'>
  <p align=center>
  <applet codebase="classes" code="applet/AppletV2.class"
        ALIGN="center"
        width=350
        height=20>
  </p>
</body>
</html>
```

INTRO.HTML

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
<html lang="es">
  <head>
    <title>:: Mi prueba Voice Vod pag 1 - prueba armando ::</title>
    <link rel="stylesheet" type="text/css" href="base.css" />
    <script type="text/javascript" src="javascript/voicevod.js"></script>
    <script type="text/javascript" src="javascript/swfobject/swfobject.js"></script>
    <script type="text/javascript">
      var flashvars = {};
      flashvars.xml = "XML/config.xml";
      flashvars.font = "font.swf";
      var attributes = {};
      attributes.wmode = "transparent";
      attributes.id = "slider";
      swfobject.embedSWF("cu3er.swf", "cu3er-container", "700", "450", "9",
"expressInstall.swf", flashvars, attributes);
  </script>
  </head>
  <body>
  </body>
</html>
```

```
</script>
  <meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
  <script type="text/javascript" src="voicevod.js"></script>

</head>

  <body background="back2.jpg" TOPMARGIN='0' LEFTMARGIN='0' MARGINWIDTH='0'
MARGINHEIGHT='0' >
    <div id="text2">
<br><br><br><br>
<center>
<div id="cu3er-container">
  <a href="http://www.adobe.com/go/getflashplayer">
    
  </a>
</div>
  </div>
</body>
</html>
```

Una vez el usuario pronuncia “iniciar servicio”, el contenido del marco dos cambia, cargando la página que contiene la guía de contenidos multimedia disponibles en sus respectivas categorías. La Figura A16, muestra la interfaz de usuario y su distribución:

Marco superior: el Applet queda a la espera de nuevos comandos, **Applet.html**.

Marco inferior: carga la página de la categoría animación, **base1.html**, con una interfaz flash tipo cover flow, **ECG.swf**, basada en el código fuente disponible en: <http://www.weberdesignlabs.com/blog/2007/09/flash-itunes-cover-flow/>

La página de **base1.html**, además de la información de los contenidos tiene un enlace a la categoría “cortometrajes”, **base2.html**.



Figura A16. Guía de Contenidos.

A continuación, se especifican los archivos utilizados en la GUI de la solución para la guía de contenidos.

BASE1.HTML

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
<html lang="es">
  <head>
    <title:: Animacion ::</title>
    <link rel="stylesheet" type="text/css" href="base.css" />
    <script type="text/javascript" src="javascript/voicevod.js"></script>
    <script type="text/javascript" src="javascript/FlashHandler.js"></script>
    <meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
    <script type="text/javascript" src="voicevod.js">
    </script>
  </head>
  <body bgcolor='#000000' TOPMARGIN='0' LEFTMARGIN='0' MARGINWIDTH='0'
MARGINHEIGHT='0'>
    <div id="content">
      <div id="header-menu">
        <ul>
          <li>
            <a class="current" href="base1.html" title="CORTOMETRAJES">ANIMACION</a>
          </li>
          <li>
            <a href="base2.html" title="CORTOMETRAJES"> CORTOMETRAJES</a>
          </li>
          <li>
            <a href="ayuda.html" title="ayuda"> AYUDA</a>
          </li>
        </ul>
      </div>
    </div>
```

```
<div id="text">
<center>
<object classid="clsid:d27cdb6e-ae6d-11cf-96b8-444553540000"
codebase="http://fpdownload.macromedia.com/pub/shockwave/cabs/flash/swflash.cab#version=8,0
,0,0" id="ECG" width=900 height=700>
<param name="allowScriptAccess" value="always" >
<param name="movie" value="ECG.swf">
<param name="quality" value="high">
<param name="play" value="false">
<param name="FlashVars" value="XMLlist=animacion.xml">
<embed swliveconnect="true" name="ECG" src="ECG.swf" FlashVars="XMLlist=animacion.xml"
quality=high bgcolor=#000000 width=900 height=530 allowScriptAccess="always"
type="application/x-shockwave-flash"
pluginspage="http://www.macromedia.com/shockwave/download/index.cgi?P1_Prod_Version=Sho
ckwaveFlash"/>
</object>
<form name="DVD_NAME" action="play.html" method="GET">
<center>
<input type="hidden" name="Data" size="20">
<input type="hidden" name="name" size="20">
<input type="hidden" value="base1.html" name="file" size="20">
<a href='javascript:ReceiveDataFromFlashMovie();'>
</A>
</center>
</form>
</div>
</div>
</body>
</html>
```

Una vez el televidente encuentra el contenido de su predilección, y pronuncia “ver película” se carga la página de visualización del contenido, play.html, como se muestra en la Figura A17. El marco superior contiene el Applet, y espera la llegada de comandos a ejecutar. El Marco inferior carga la pagina de reproducción (Play.html) e inicia la reproducción del contenido multimedia mediante el complemento “VLC Multimedia Plugin” el cual permite el control de la película mediante funciones JavaScript.

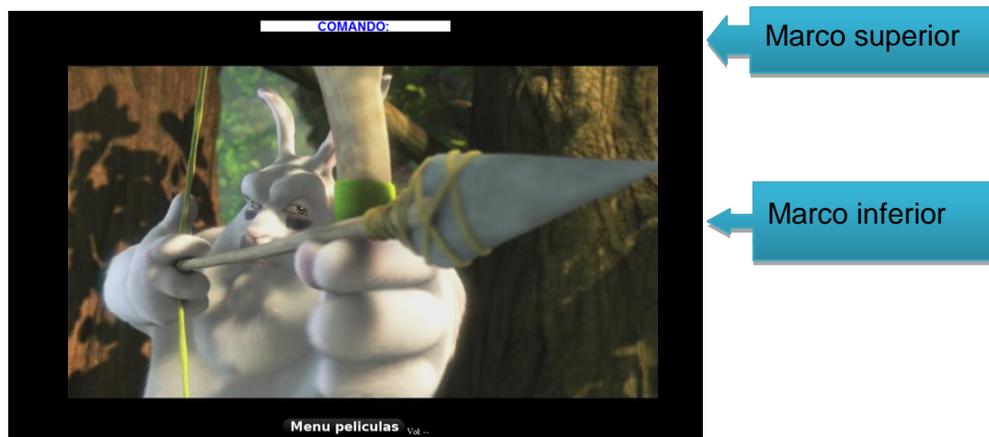


Figura A17. Reproducción del contenido.

PLAY.HTML

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
"http://www.w3.org/TR/html4/strict.dtd">
<html>
  <head>
    <script type="text/javascript" src="javascript/voicevod.js">
    </script>
    <SCRIPT type="text/ecmascript">
    <!--
    function PreviousPage(){
      onStop();
      this.location.href = file;
    }
    function getVLC(name){
    if (window.document[name]) {
      return window.document[name];
    }
    if (navigator.appName.indexOf("Microsoft Internet") == -1) {
      if (document.embeds && document.embeds[name])
        return document.embeds[name];
    }
    else // if (navigator.appName.indexOf("Microsoft Internet")!=-1)
    {
      return document.getElementById(name);
    }
    }
    function onVLCPluginReady(){
      updateVolume(0);
    };
    //-->
    </SCRIPT>
    <meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
    <title>movie</title>
  </head>
  <body " bgcolor="#000000" text="#FFFFFF" >
    <div align="center">
      <embed type="application/x-vlc-plugin" pluginspage="http://www.videolan.org"
      version="VideoLAN.VLCPlugin.2" id="vlc" width="940" height="630">
      <script language="Javascript">
        var vlc = document.getElementById("vlc");
        function getQueryVariable(variable){
          var query = window.location.search.substring(1);
          var vars = query.split("&");
          for (var i = 0; i < vars.length; i++) {
            var pair = vars[i].split("=");
            if (pair[0] == variable) {
              return pair[1];
            }
          }
          alert('Query Variable ' + variable + ' not found');
        }
        var source = getQueryVariable("Data");
        var file = getQueryVariable("file");
```

```
var name = getQueryVariable("name");
s = new String(source);
n = new String(name);
s = s.replace(/%5C/g, "\\");
s = s.replace(/%2F/g, "/");//agregado por miiiiiiiiiiiiiiiiii
s = s.replace(/%3A/g, ":");//es necesario
n = n.replace(/#/g, "%");//es necesario
file = file.replace(/#/g, "%");
//document.getElementById("targetTextField").value = s;//era s
function PreviousPage(){
    this.location.href = file;
}
var id = vlc.playlist.add(s);
vlc.playlist.playItem(id);
    document.getElementById("volumeTextField").innerHTML = "50 %"
    updateVolume(+10);
var seek_time = 1000 * 50;
var volume_number = 10;

function onPause(){
    vlc.playlist.togglePause();
}
    function onStop(){
        //getVLC("vlc").playlist.stop();
    vlc.playlist.stop();
}
function onPlay(){
    //vlc.input.rate = 1.0
    vlc.playlist.play();
}
function onFullscreen(){
    vlc.video.toggleFullscreen();
}
function onMute(){
    vlc.audio.toggleMute();
    document.getElementById("volumeTextField").innerHTML =
    vlc.audio.volume + "%";
}
function onSeek_forward(){
    if ((vlc.input.time + seek_time) < vlc.input.length) {
        vlc.input.time = vlc.input.time + seek_time;
    }
    else {
        vlc.input.time = 0;
    }
}
function onSeek_backward(){
    if ((vlc.input.time - seek_time) >= 0) {
        vlc.input.time -= seek_time;
    }
    else {
        vlc.input.time = 0
    }
}
function fastForward(){
```

```
        vlc.input.rate = 8.0
    }
    function fastRewind(){
        vlc.input.rate = -2.0
    }
    //funciones del otro
        function updateVolume(deltaVol){//llamar a esta funcion pasandole +10 o -10
    var vlc = getVLC("vlc");
    vlc.audio.volume += deltaVol;
    var div = "2";
    total = 'vlc.audio.volume+"%" / 'div';
    document.getElementById("volumeTextField").innerHTML = vlc.audio.volume + "%";
    };
    var aspectRatio = "default";
    function doChangeAspectRatio(arValue){//a esta se le pasa "16:9" o "4:3"
    var vlc = getVLC("vlc");
    if (vlc.input.state && vlc.input.hasVout) {
        vlc.video.aspectRatio = arValue;
    }
    aspectRatio = arValue;
    };
    function doPlaySlower(){
    var vlc = getVLC("vlc");
    vlc.input.rate = vlc.input.rate / 2;
    };
    function doPlayFaster(){
    var vlc = getVLC("vlc");
    vlc.input.rate = vlc.input.rate * 2;
    };
</script>
<center>
    <a href='javascript:PreviousPage();'></A>
Vol: <SPAN id="volumeTextField" style="text-align: center;">--</SPAN>
</center>
</body>
</html>
```

Con el archivo **play.html**, se terminará la implementación de la GUI de la solución. A continuación se explica la interfaz entre la GUI y el módulo de reconocimiento del habla.

4.2 Interfaz con el Motor de Reconocimiento

Esta interfaz se realiza mediante un Applet que crea un servidor socket para comunicarse con el AGI de Asterisk. Dicho Applet reside en el primer marco de la interfaz gráfica HTML.

AppletV2.java

```
package applet;
import java.io.*;
import java.net.*;
import java.applet.*;
```

```
import java.awt.*;
import javax.swing.JOptionPane;
import netscape.javascript.JSObject;
publicclass AppletV2 extends Applet implements Runnable {
    JSObject jsc;
    staticfinalint MYPORT = 6700;
    protected DataInputStream in;
    ServerSocket listen_socket;
    private Label rta;
    Thread hilo = null;
    Font f = newFont("Arial", Font.BOLD, 24);
    @Override
    publicvoid init() {
        initComponents();
        init_javascript_link();
    }
    publicvoid start() {
        try {
            listen_socket = newServerSocket(MYPORT);

        } catch (IOException e) {
            System.err.println(e);
        }
        if (hilo == null) {
            hilo = new Thread(this);
            hilo.start();
        }
    }
    publicvoid stop() {
        hilo = null;
    }
    publicvoid run() {
        while (true) {
            try {
                Socket client_socket = listen_socket.accept();
                Connection c = newConnection(client_socket);
            } catch (IOException e) {
                System.err.println(e);
            }
        }
    }

    /*****FINHILO*****/
    privatevoid initComponents() {
        rta = new Label();
        rta.setAlignment(Label.CENTER);
        rta.setForeground(Color.BLUE);
        rta.setFont(f);
        rta.setText("COMANDO: ");
        setLayout(new GridLayout(1, 1));
        add(rta);
    }
    class Connection extends Thread {
        protected Socket client;
        protected DataInputStream in;
    }
}
```

```
public Connection(Socket client_socket) {
    client = client_socket;
    try {
        in = new DataInputStream(client.getInputStream());
    } catch (IOException e) {
        System.err.println(e);
    }
    try {
        client.close();
    } catch (IOException e2) {
    }
    ;
    return;
}
this.start();
}
public void run() {
    String comando;
    try {
        comando = in.readLine();
        rta.setText("COMANDO: " + comando);
        Object args[] = {(Object) comando};
        llamar_funcion_javascript_parametros("ProcessResult", args);
        repaint();
        System.out.println("el comando recibido fue:" + comando);
    } catch (IOException e) {
    } finally {
    }
    try {
        client.close();
    } catch (IOException e) {
    }
    ;
    }
}
}
public void init_javascript_link() {
    this.jsc = JScriptEngine.getEngine(this);
}
public void llamar_funcion_javascript(String s) {
    try {
        this.jsc.call(s, null);
    } catch (Exception e) {
    }
    JOptionPane.showMessageDialog(this, s);
}
}
public void llamar_funcion_javascript_parametros(String s, Object[] para) {
    try {
        this.jsc.call(s, para);
    } catch (Exception e) {
    }
    JOptionPane.showMessageDialog(this, s);
} } }
```

Con la implementación de la interfaz entre la GUI, se integra todo el sistema, se concluye la tercera y última actividad de la fase de configuración, instalación e implementación de la solución. Asimismo, con esta etapa finaliza el anexo A.

Análisis del Uso de Interactividad basada en Voz para el manejo del Servicio de Video bajo Demanda



ANEXO B

LAURA GÓMEZ FIGUEROA
JESÚS ARMANDO SANDOVAL LEON

UNIVERSIDAD DEL CAUCA
FACULTAD DE INGENIERIA ELECTRONICA Y TELECOMUNICACIONES
DEPARTAMENTO DE TELECOMUNICACIONES
GRUPO I+D EN NUEVAS TECNOLOGÍAS EN TELECOMUNICACIONES

ANEXO B. CUESTIONARIOS DE PERCEPCIÓN DEL USUARIO Y RESULTADOS OBTENIDOS

Dentro de este documento, se presentan los Cuestionarios de Percepción del Servicio, para el control remoto y para la voz. Así mismo, se presentan los resultados obtenidos luego de la aplicación de dichos cuestionarios.

1. DESCRIPCIÓN DE LA POBLACIÓN

La Tabla 1, contiene la descripción de de los usuarios que evaluaron la solución. Dentro de ésta se describe el género, ocupación y edad de los mismos.

Tabla B 1. Información sobre la Población

Usuario	Sexo	Edad	Ocupación
1	F	9	Estudiante.
2	M	52	Docente.
3	M	10	Estudiante.
4	M	11	Estudiante.
5	F	24	Estudiante Universitario.
6	F	43	Operaria de aseo.
7	F	43	Operaria de aseo.
8	F	27	Operaria de aseo.
9	M	29	Estudiante Universitario.
10	M	21	Estudiante Universitario.
11	M	19	Estudiante Universitario.
12	M	51	Técnico en electrónica
13	M	37	Docente
14	F	37	Docente
15	M	22	Estudiante Universitario.
16	M	24	Estudiante Universitario.
17	M	24	Estudiante Universitario.
18	M	22	Estudiante Universitario.
19	M	24	Estudiante Universitario.
20	M	24	Estudiante Universitario.
21	F	20	Estudiante Universitario.

2. RESULTADOS DE LA APLICACIÓN DEL CUESTIONARIO DE PERCEPCIÓN DEL SERVICIO PARA EL CONTROL REMOTO

En esta sección se presenta el CPS para el control remoto, y los resultados obtenidos luego de la aplicación a los usuarios del sistema.

2.1 Cuestionario de la Percepción del Servicio para el Control Remoto

2.2 Resultados Obtenidos

La Tabla B2, contiene las respuestas obtenidas luego de la aplicación del CPS para el aspecto “atractivo” controlado por control remoto. Ésta relaciona el número de personas que responden de forma similar a cada afirmación.

Tabla B2. Resultados del Aspecto “Atractivo” del Sistema Controlado por Control Remoto

Respuesta	Afirmación				
	1	2	3	4	5
Muy de acuerdo	33,33%	0,00%	33,33%	0,00%	23,81%
De acuerdo	52,38%	4,76%	47,62%	0,00%	57,14%
Indeciso	9,52%	14,29%	14,29%	9,52%	9,52%
en desacuerdo	4,76%	52,38%	4,76%	52,38%	4,76%
Muy en desacuerdo	0,00%	28,57%	0,00%	38,10%	4,76%

La Tabla B3, contiene las respuestas obtenidas luego de la aplicación del CPS para el aspecto “control” para el sistema con control remoto. Ésta relaciona el número de personas que responden de forma similar a cada afirmación.

Tabla B3. Resultados del Aspecto “Control” del Sistema con Control Remoto

Respuesta	Afirmación		
	6	7	8
Muy de acuerdo	4,76%	23,81%	0,00%
De acuerdo	33,33%	33,33%	4,76%
Indeciso	4,76%	23,81%	23,81%
en desacuerdo	47,62%	19,05%	42,86%
Muy en desacuerdo	9,52%	0,00%	28,57%

La Tabla B3, contiene las respuestas obtenidas luego de la aplicación del CPS para el aspecto “eficiencia” para el sistema con control remoto.

Tabla B4. Resultados del Aspecto “Eficiencia” del Sistema Controlado por Control Remoto

Respuesta	Afirmación				
	9	10	11	12	13
Muy de acuerdo	4,76%	19,05%	14,29%	0,00%	0,00%
De acuerdo	23,81%	57,14%	57,14%	4,76%	0,00%
Indeciso	4,76%	4,76%	19,05%	19,05%	19,05%
en desacuerdo	52,38%	9,52%	4,76%	57,14%	47,62%
Muy en desacuerdo	14,29%	9,52%	4,76%	19,05%	33,33%

La Tabla B5, contiene las respuestas obtenidas luego de la aplicación del CPS para el aspecto “utilidad” para el sistema con control remoto.

Tabla B5. Resultados del Aspecto “Utilidad” del Sistema Controlado por Control Remoto

Respuesta	Afirmación	
	14	15
Muy de acuerdo	23,81%	19,05%
De acuerdo	33,33%	23,81%
Indeciso	33,33%	19,05%
en desacuerdo	4,76%	38,10%
Muy en desacuerdo	4,76%	0,00%

La Tabla B6, contiene las respuestas obtenidas luego de la aplicación del CPS para el aspecto “facilidad de aprendizaje” para el sistema con control remoto.

Tabla B6. Resultados del Aspecto “Facilidad de Aprendizaje” del Sistema Controlado por Control Remoto

Respuesta	Afirmación					
	16	17	18	19	20	21
Muy de acuerdo	4,76%	4,76%	38,10%	19,05%	19,05%	4,76%
De acuerdo	4,76%	14,29%	42,86%	66,67%	33,33%	42,86%
Indeciso	9,52%	0,00%	4,76%	4,76%	9,52%	9,52%
en desacuerdo	57,14%	47,62%	14,29%	4,76%	38,10%	28,57%
Muy en desacuerdo	23,81%	33,33%	0,00%	4,76%	0,00%	14,29%

3. RESULTADOS DEL CUESTIONARIO DE PERCEPCIÓN DEL SERVICIO PARA LA INTERACCIÓN POR VOZ

En esta sección se presenta el CPS para el sistema controlado por voz, y los resultados obtenidos luego de la aplicación a los usuarios del sistema.

3.1 Cuestionario de la Percepción del Servicio para el Sistema Controlado por Voz

En la siguiente página se muestra el cuestionario aplicado a la población de la Tabla B1. Éste adiciona 4 preguntas abiertas, cuyos resultados se presentan al final de esta sección.

3.2 Resultados Obtenidos

La Tabla B7, contiene las respuestas obtenidas luego de la aplicación del CPS para el aspecto “atractivo” controlado por voz.

Tabla B7. Resultados del Aspecto “Atractivo” del Sistema Controlado por Voz.

Respuesta	Afirmación				
	1	2	3	4	5
Muy de acuerdo	42,86%	0,00%	33,33%	0,00%	66,67%
De acuerdo	57,14%	0,00%	66,67%	0,00%	28,57%
Indeciso	0,00%	0,00%	0,00%	9,52%	4,76%
en desacuerdo	0,00%	66,67%	0,00%	38,10%	0,00%
Muy en desacuerdo	0,00%	33,33%	0,00%	52,38%	0,00%

La Tabla B8, contiene las respuestas obtenidas luego de la aplicación del CPS para el aspecto “control” para el sistema controlada por voz.

Tabla B8. Resultados del Aspecto “Control” del Sistema controlado por Voz.

Respuesta	Afirmación		
	6	7	8
Muy de acuerdo	4,76%	19,05%	0,00%
De acuerdo	28,57%	61,90%	0,00%
Indeciso	33,33%	14,29%	9,52%
en desacuerdo	23,81%	4,76%	71,43%
Muy en desacuerdo	9,52%	0,00%	19,05%

La Tabla B9, contiene las respuestas obtenidas luego de la aplicación del CPS para el aspecto “eficiencia” para el sistema controlado por voz.

Tabla B9. Resultados del Aspecto “Eficiencia” del Sistema Controlado por Voz.

Respuesta	Afirmación				
	9	10	11	12	13
Muy de acuerdo	0,00%	19,05%	19,05%	0,00%	4,76%
De acuerdo	9,52%	61,90%	33,33%	33,33%	0,00%
Indeciso	14,29%	14,29%	23,81%	14,29%	9,52%
en desacuerdo	57,14%	4,76%	19,05%	42,86%	76,19%
Muy en desacuerdo	19,05%	0,00%	4,76%	9,52%	9,52%

La Tabla B10, contiene las respuestas obtenidas luego de la aplicación del CPS para el aspecto “utilidad” para el sistema controlado por voz.

Tabla B10. Resultados del Aspecto “Utilidad” del Sistema Controlado por Voz.

Respuesta	Afirmación	
	14	15
Muy de acuerdo	33,33%	14,29%
De acuerdo	52,38%	47,62%
Indeciso	9,52%	19,05%
en desacuerdo	4,76%	19,05%
Muy en desacuerdo	0,00%	0,00%

La Tabla B11, contiene las respuestas obtenidas luego de la aplicación del CPS para el aspecto “facilidad de aprendizaje” para el sistema controlado por voz.

Tabla B11. Resultados del Aspecto “Facilidad de Aprendizaje” del Sistema Controlado por Voz.

Respuesta	Afirmación					
	16	17	18	19	20	21
Muy de acuerdo	9,52%	0,00%	42,86%	42,86%	19,05%	4,76%
De acuerdo	0,00%	0,00%	52,38%	52,38%	28,57%	57,14%
Indeciso	0,00%	4,76%	4,76%	0,00%	28,57%	4,76%
en desacuerdo	61,90%	76,19%	0,00%	0,00%	23,81%	23,81%
Muy en desacuerdo	28,57%	19,05%	0,00%	4,76%	0,00%	9,52%

3.3 Resultados de las Preguntas Abiertas

La Tabla B12, Contiene las respuestas a las preguntas contenidas en el CPS para el sistema controlado por voz. Las preguntas son:

1. ¿Qué expectativas tenía de la prueba del servicio? ¿Las cumplió? Sí__ No__ ¿Por qué?
2. En general ¿Cree que el control por voz para el servicio de Video bajo Demanda es bueno? Sí__ No__ ¿Por qué?
3. ¿Cuál cree que es el mayor defecto/problema del sistema?
4. Entre el servicio de Video bajo Demanda controlado por control remoto y el controlado por voz. ¿Cuál escogerías? ¿Por qué?

Tabla B 12. Respuestas de los Usuarios a las Preguntas Abiertas.

Usuario	Expectativas.	Creencia de si el control por voz es bueno	Defecto del sistema	Cual escogerías
1	No tenia	Si, porque no hay que buscar el control remoto en caso de que se pierda.	No ejecuta las órdenes rápido.	El de voz, porque solo hay que hablar(facilidad)
2	Si cumplió con las expectativas, ya que el menú de voz evita errores en pérdida de control, además que tiene muchas aplicaciones en forma interactiva	Si, facilita el manejo de la televisión.	no	El de voz, porque tiene menos fallas.
3	Si	Si, ya que con la voz es mejor que presionar botones en el control remoto.	Imposibilidad de ejecutar comandos cuando se tenga dolor de garganta.	El de voz es mas efectivo.
4	Si	Si	La lentitud	La voz, ya que es mas simpe.
5	Sí, porque el sistema respondió a los comandos de voz para los que estaba programado. Sería interesante que se le programaran nuevos o más comandos a los que pudiera responder.	Si, porque los comandos pueden ser diferentes y aun así cumplir la función, es decir, hace lo mismo al decir aumentar volumen que al decir volumen, además los comandos son en español, no en ingles como las teclas del control del DVD. Además es un sistema original.	De pronto que respondiera al comando de voz de la persona que este viendo la película, igual como respondería al de otras personas alrededor, es decir no seria muy personalizado.	El controlado por voz, debido a que los comandos son en nuestro idioma, y no tienes que aprender a usarlo cada ves que cambias de DVD como me sucede con el control de cada DVD que sale al mercado.
6	No, porque al principio no entendí	Si, porque es de fácil manejo	Ninguno	Voz, por mejor entendimiento del manejo.
7	Si, porque le parecieron fáciles	Si, porque no hay que estar pendiente del control.	No le escucho el mando de voz.	Control de voz, porque es mas fácil.
8	Si, pensaba que era difícil pero no lo era.	Si, porque no toca hablar muy fuerte.	Es un poco lento cuando se le da la orden con la voz.	Los dos, porque son muy fáciles.
9	Si, porque funciona como un DVD, pero sin el control.	Si, porque me da libertad de manejo.	La velocidad de respuesta.	El controlado por voz, porque puedo tener un micrófono y dar la orden son ocupar las manos y desde un sitio cualquiera.
10	Si, al encontrarme	Aumenta la	Ninguno	Voz, ya que permite

Análisis del Uso de Interactividad basada en Voz para el manejo del Servicio de Video bajo Demanda

	con un sistema novedoso y práctico a la vez que minimicé el uso de controles manuales. Cumplió completamente mis expectativas.	interacción con el sistema y anima el uso de este.		una rápida interacción sin necesidad de buscar y buscar botones para oprimir.
11	Las expectativas que tuve siempre sobre este servicio siempre fueron las mejores, ya estando frente a él e interactuando me di cuenta que es fantástico y que además es fácil de usar.	Sí, porque facilita varias operaciones, no tengo que estar moviéndome de un lado a otro tratando de insertar DVD, darle volumen al equipo, etc. Brinda mucha más comodidad al usuario.	Tal vez el reconocimiento de las palabras adecuadas, debido a que no en todas las regiones hablamos de la misma manera.	Voz, ya que facilita muchas cosas y es más fácil de usar, esto garantiza a la persona o usuario sentirse mucho más cómodo a la hora de disfrutar una película, sin tener que levantarse de su puesto.
12	Si, al ser un sistema novedoso y nuevo, que bueno tener acceso a él y saber utilizarlo.	Sí, porque: 1-elimina el aparato "control remoto" 2-me permite interactividad con el sistema.	El grado de precisión de las órdenes con respecto a las acciones.	Voz, por: Interactividad Comodidad Novedad
13	Si	Si, facilita las cosas	En general no veo mayor problema	Por voz. (como ocurriría en compañía de otros usuarios?)
14	Si	Si	Es un poco lento	Voz
15	Si, en general esperaba que el sistema respondiera adecuada y eficientemente a los comandos que le entregaba.	Si, porque es intuitivo.	Algunas veces(muy pocas) no reconoce el comando de voz que se le entrega.	Voz, ya que, a pesar de que me encuentro habituado al control remoto, considero que el primero es más usable.
16	NO, Esperaba que tuviera más comandos de manejo y una ayuda donde explicara los existentes	Si, las opciones de manejo por lo general son pocas y repetitivas, podría implementarse.	La lentitud de acción.	Control remoto, pero el piloto manejado por voz le pareció muy creativo.
17	Si, responde a los comandos de voz, siempre y cuando se digan los correctos, es amigable y tiene una buena presentación.	Si, permite un manejo más intuitivo del dispositivo, evitando el uso de un control saturado de botones, que muchas veces resultan inservibles.	La necesidad de estar cerca del micrófono.	Voz, ya que brinda mayor comodidad al usuario y evita el uso de dispositivos muchas veces poco comprensibles como el control.
18	Si, porque reconoció todos los	Si, es mucho más fácil y un poco más	La latencia.	Dependiendo de latencia (común)

Análisis del Uso de Interactividad basada en Voz para el manejo del Servicio de Video bajo Demanda

	comandos que se dijo.	universal que el control remoto.		para ambos) se escoge la voz, si hay problemas de voz (afonía) pasa al control.
19	Si, es interesante ver la diferencia de usar tanto la voz, como el control, en el mismo sistema.	Si, siempre y cuando la respuesta del sistema sea eficaz y eficiente.	El ruido	Esta respuesta es relativa, depende del contexto. Es decir, en un entorno con un nivel de ruido considerable, es conveniente usar el control
20	Si, porque es fácil de usar, efectivo y diferente de los sistemas de control.	Si, porque es una forma diferente de poder controlar este tipo de servicio, entretenida, además cumple con los comandos que se ordenan.	Que a veces reconoce palabras que no están en los comandos, y realiza funciones inesperadas.	El controlado por voz, porque se evita de tener un control en la mano con botones que no se utilizan, además si estás viendo VoD con la Luz apagada se puede controlar fácilmente el sistema.
21	Sí, es un sistema que nos facilita una mayor interactividad, gracias a las posibilidades de continuar o no con otra actividad que se esté desarrollando.	Si, al brindar facilidades para la continuación de otras tareas el sistema brinda una comodidad perfecta para el usuario.	No sabe/ no responde	Voz, por facilidad y comodidad de uso. Porque nos ayuda a controlar con mayor facilidad, lo que los pequeños ven en casa, quienes además podrán aprender rápidamente el uso del sistema.

Con las respuestas de los usuarios a las preguntas abiertas culmina este anexo.

Análisis del Uso de Interactividad basada en Voz para el manejo del Servicio de Video bajo Demanda



ANEXO C

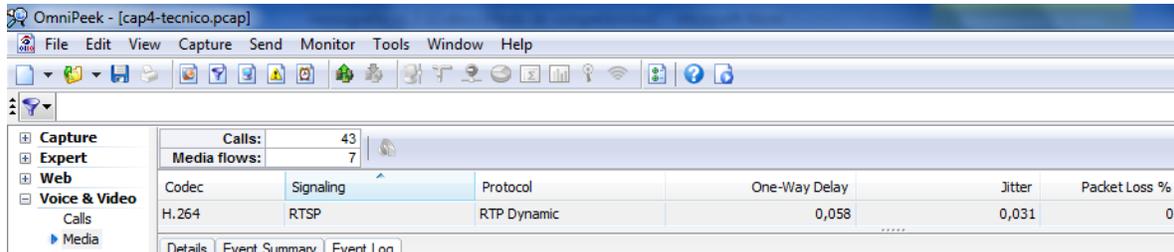
LAURA GÓMEZ FIGUEROA
JESÚS ARMANDO SANDOVAL LEON

UNIVERSIDAD DEL CAUCA
FACULTAD DE INGENIERIA ELECTRONICA Y TELECOMUNICACIONES
DEPARTAMENTO DE TELECOMUNICACIONES
GRUPO I+D EN NUEVAS TECNOLOGÍAS EN TELECOMUNICACIONES

ANEXO C. ANALISIS TECNICO

1. FLUJO DE VIDEO

En la Figura C.1 se muestra el cálculo de los valores de retardo, Jitter y pérdida de paquetes, calculados mediante la herramienta OmniPeek.



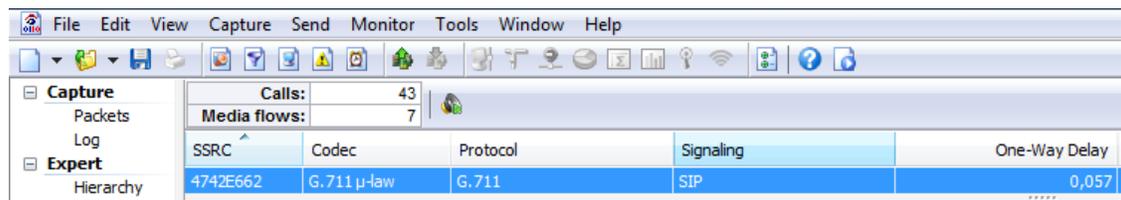
The screenshot shows the OmniPeek interface with a table of video flow performance metrics. The table has columns for Codec, Signaling, Protocol, One-Way Delay, Jitter, and Packet Loss %.

Codec	Signaling	Protocol	One-Way Delay	Jitter	Packet Loss %
H.264	RTSP	RTP Dynamic	0,058	0,031	0

Figura C1. Factores de desempeño del flujo de video.

2. COMUNICACIÓN VoIP

Los valores de retardo, Jitter y pérdida de paquetes se obtuvieron mediante los programa OmniPeek y Cace Pilot. Las Figuras C2 y C3, muestran las capturas de pantallas de los parámetros de evaluación de desempeño, y las Figuras C4, C5 y C6, muestran el cálculo del Jitter, pérdida de paquetes y ancho de banda en función del tiempo.



The screenshot shows the OmniPeek interface with a table of VoIP performance metrics. The table has columns for SSRC, Codec, Protocol, Signaling, and One-Way Delay.

SSRC	Codec	Protocol	Signaling	One-Way Delay
4742E662	G.711 μ -law	G.711	SIP	0,057

Figura C 2. Retardo.

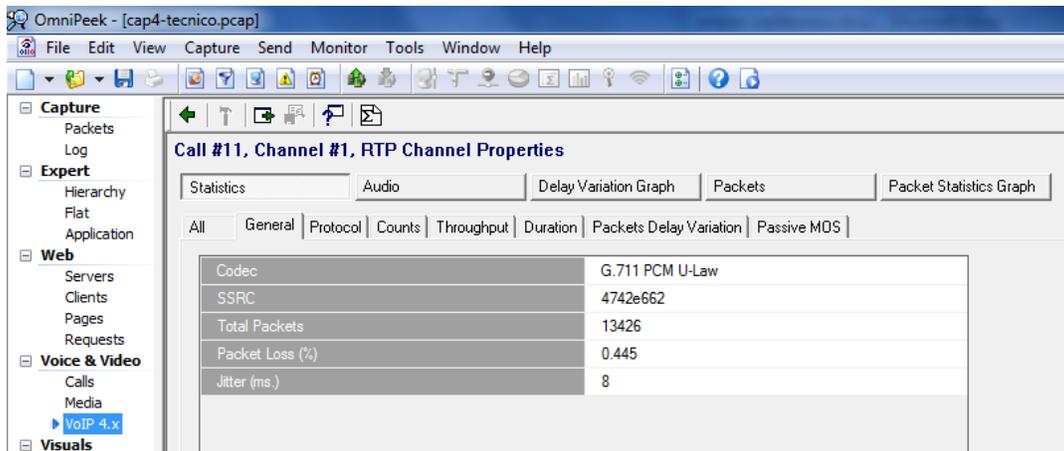


Figura C3. Jitter y pérdida de paquetes.



Figura C4. Ancho de banda consumido en función del tiempo.



Figura C5. Jitter promedio en función del tiempo.



Figura C6. Pérdida de paquetes promedio en función del tiempo

3. TIEMPO FLUJO TCP

Para el cálculo de este tiempo se procedió a tomar el tiempo de 30 comandos enviados vía TCP. La información de la Tabla C1, se obtuvo del programa OmniPeek.

Tabla C1. Tiempo de transmisión promedio del comando de control.

PUERTOS	TIEMPO DE FLUJO (Seg)
44303<->6700	0,004
44273<->6700	0,005
44305<->6700	0,005
44275<->6700	0,005
44263<->6700	0,005
44299<->6700	0,005
44313<->6700	0,006
44257<->6700	0,006
44277<->6700	0,006
44245<->6700	0,008
44243<->6700	0,008
38819<->6700	0,008
44301<->6700	0,008
44281<->6700	0,009
44267<->6700	0,009
44241<->6700	0,009
44271<->6700	0,009
44239<->6700	0,01
38821<->6700	0,01
44291<->6700	0,011
44283<->6700	0,011
44249<->6700	0,012
44247<->6700	0,013
38817<->6700	0,013
44297<->6700	0,014
44287<->6700	0,014
44285<->6700	0,015
44295<->6700	0,017
44255<->6700	0,022
44311<->6700	0,023
PROMEDIO	0,01

4. RETARDO DE PROCESAMIENTO DEL MOTOR DE RECONOCIMIENTO DE VOZ

El retardo de procesamiento del motor de reconocimiento se calcula mediante WireShark, quien permite la visualización del flujo de audio de una llamada, tal y como se muestra en la Figura C7.

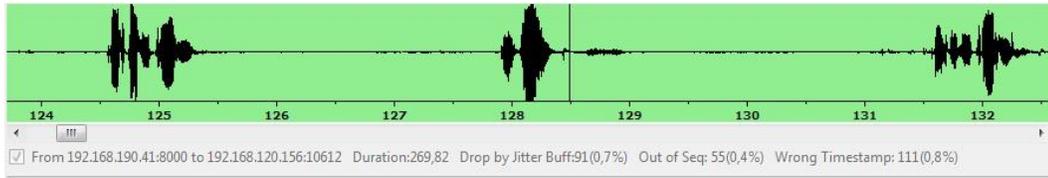


Figura C7. Audio decodificado de la comunicación VoIP.

En la Figura C7, se observan tres variaciones de audio (comandos) a lo largo del tiempo. Para calcular en promedio cuando se demora en entregar Lumenvox el resultado del reconocimiento al AGI, se procede por a analizar en qué punto termina un determinado comando, por ejemplo, la variación ubicada en medio en la figura, aproximadamente termina en 128,5 segundos. Con este tiempo, lo siguiente es buscar el paquete que tiene aproximadamente ese tiempo. En la Figura C8, se observa que el paquete número 2663, esta aproximadamente en ese tiempo. De aquí lo siguiente es moverse hacia abajo en la lista de paquetes hasta encontrar el flujo TCP que transmite el mando reconocido.

Una vez se encuentre en el flujo TCP simplemente habrán de restar el tiempo entre el paquete inmediatamente anterior al inicio del flujo TCP y el tiempo del paquete 2633. Esta diferencia será el tiempo aproximado de procesamiento del motor de reconocimiento. De esta manera, el tiempo que se demora Lumenvox en Procesar y entregar el comando reconocido es aproximadamente 1205,218 ms. Este mismo proceso se repite para varios comandos, de manera que se obtiene un promedio de: 1260,138 ms.

2662	128.489838	192.168.190.41	192.168.120.156	RTP	PT=ITU-T G.711 PCMU,
2663	128.509906	192.168.190.41	192.168.120.156	RTP	PT=ITU-T G.711 PCMU,
2664	128.530037	192.168.190.41	192.168.120.156	RTP	PT=ITU-T G.711 PCMU,
2665	128.550250	192.168.190.41	192.168.120.156	RTP	PT=ITU-T G.711 PCMU,
2725	129.675046	192.168.190.41	192.168.120.156	RTP	PT=ITU-T G.711 PCMU,
2726	129.695080	192.168.190.41	192.168.120.156	RTP	PT=ITU-T G.711 PCMU,
2727	129.715124	192.168.190.41	192.168.120.156	RTP	PT=ITU-T G.711 PCMU,
2728	129.717197	192.168.120.156	192.168.190.100	TCP	44241 > 6700 [SYN] Seq=
2729	129.718105	192.168.190.100	192.168.120.156	TCP	6700 > 44241 [SYN, ACK]
2730	129.718115	192.168.120.156	192.168.190.100	TCP	44241 > 6700 [ACK] Seq=
2731	129.718191	192.168.120.156	192.168.190.100	TCP	44241 > 6700 [PSH, ACK]
2732	129.718231	192.168.120.156	192.168.190.100	TCP	44241 > 6700 [FIN, PS
2733	129.719207	192.168.120.156	192.168.190.41	RTP	PT=ITU-T G.711 PCMU,
2734	129.719636	192.168.190.100	192.168.120.156	TCP	6700 > 44241 [ACK] Seq=
2735	129.724235	192.168.190.100	192.168.120.156	HTTP	GET /~/laura/avuda.htm

Figura C8. Calculo del tiempo de procesamiento del comando.

Con el cálculo del tiempo de procesamiento del motor de reconocimiento concluye este anexo.