

**Esquema de Operación de
Sistemas de Reconocimiento Automático
de Voz sobre VoWLAN
para el idioma Español**

Tesis de Maestría en Ingeniería
Área Electrónica y Telecomunicaciones

Jaime Andrés Gaviria Molano

Director: Ing. PhD. Carlos Alberto Gaviria López

**Universidad del Cauca
Facultad de Ingeniería Electrónica y Telecomunicaciones
Popayán, Mayo de 2009**

TABLA DE CONTENIDO

Tabla de contenido.....	2
Resumen.....	2
1 INTRODUCCION	2
1.1 Justificación y antecedentes.....	2
1.2 metodología.....	5
1.3 alcance y limitaciones.....	5
1.4 organización.....	6
2 VOWLAN	7
2.1 VoIP	7
2.1.1 Protocolos VoIP.....	7
2.1.1.1 Protocolos de Señalización	8
2.1.1.1.1 H.323	8
2.1.1.1.2 SIP	8
2.1.1.2 Protocolos de Transporte	9
2.1.1.2.1 RTP	9
2.1.1.2.2 CRTP	9
2.1.1.3 Protocolos de Enrutamiento	10
2.1.1.3.1 RIP.....	10
2.1.1.3.2 OSPF.....	10
2.1.1.3.2.1 SPF.....	10
2.1.1.3.3 BGP.....	11
2.1.1.3.4 RSVP.....	11
2.1.2 Procesamiento y empaquetamiento de Voz sobre redes IP.....	12
2.1.2.1 Digitalización de la voz	12
2.1.2.1.1 PCM	12
2.1.2.1.1.1 Muestreo.....	12
2.1.2.1.1.2 Compansión.....	13
2.1.2.1.1.3 Cuantificación.....	13
2.1.2.1.1.4 Codificación.....	13
2.1.2.2 Códecs	13
G.711.....	14
G.728.....	14

G.723.1.....	15
G.729.....	15
2.1.2.3 Esquemas de Paquetización.....	15
2.1.3 Medición de la calidad de la voz.....	16
2.1.3.1 MOS.....	16
2.1.3.2 PSQM.....	16
2.1.4 Elementos básicos de un sistema VoIP.....	17
2.1.4.1 SoftSwitch	17
2.1.4.2 Terminales.....	17
2.1.4.2.1 Teléfonos IP.....	17
2.1.4.2.2 ATAs.....	18
2.1.4.2.3 Softphones.....	18
2.2 Wlan.....	18
2.2.1 Acceso al medio.....	19
2.2.1.1 DCF.....	19
2.2.2 Nivel Físico.....	20
2.2.2.1 802.11b.....	20
2.2.2.2 802.11g.....	20
2.2.3 Operación de la red en modo infraestructura.....	21
2.2.3.1 Escaneo.....	22
Escaneo Pasivo.....	22
Escaneo Activo.....	22
2.2.3.2 Conexión a la red.....	23
2.2.3.3 Transferencia de Datos.....	23
2.2.3.4 Roaming.....	23
2.3 VoWLAN	24
2.3.1 Consideraciones sobre WLAN.....	24
2.3.1.1 Cobertura.....	24
2.3.1.2 Retardo.....	26
2.3.1.3 Interferencia.....	26
2.3.1.4 Seguridad	27
3 SISTEMAS DE RECONOCIMIENTO DE VOZ.....	28
3.1 Funcionamiento de un sistema de Reconocimiento.....	28
3.1.1 Extracción de Características (Descriptores).....	28
3.1.1.1 Descriptores de Audio.....	29
3.1.1.1.1 Extracción de Características de Señales de Audio Mediante Coeficientes Cepstrales (MFCC Mel Frequency Cepstral Coefficients) 30	
3.1.2 Clasificador.....	31
3.1.2.1 Tipos comunes de clasificadores.....	31
3.1.2.1.1 Clasificador por distancias euclideas.....	31
3.1.2.1.2 Clasificador por vecino más cercano.....	32
3.1.2.1.3 Clasificador por los k-vecinos más cercanos.....	33
3.2 Reconocimiento del habla.....	33
3.2.1 Clasificación de sistemas de reconocimiento del habla.....	34
RECONOCIMIENTO DE PALABRAS AISLADAS.	34

RECONOCIMIENTO DE PALABRAS CONECTADAS.....	34
REC. PAL. AISLADAS O CONECTADAS:	34
RECONOCIMIENTO/COMPRESION DEL HABLA CONTINUA.....	34
3.2.2 Técnicas de reconocimiento del habla.....	35
PLANTILLAS (Alineamiento Temporal No Lineal).....	35
PROBABILISTICOS (Modelos De Markov Ocultos).....	35
ESTRUCTURALES (Redes Finitas/Gramáticas).....	35
CONEXIONISTAS (Redes Neuronales Artificiales).....	35
3.2.3 Modelos Ocultos de Markov en el reconocimiento del habla.....	35
3.2.4 Modelos del lenguaje.....	37
3.2.4.1 Modelo n-gramas.....	37
4 Modelos de experimentación y simulación.....	39
4.1 Mecanismos Para Proveer Calidad De Servicio Sobre 802.11	40
4.1.1 Códecs de Voz.....	41
4.1.1.1 Retardo.....	42
4.1.1.2 Calidad de la señal.....	42
4.1.1.3 Esquemas de Paquetización.....	43
4.1.1.3.1 Transmisión de varias muestras por paquete.....	43
4.1.1.3.2 Compresión de Cabeceras:.....	44
4.1.1.4 Jitter y Jitter Buffer.....	44
4.1.1.5 Consideraciones de QoS.....	44
4.2 Construcción De Un Sistema De VoIP Sobre WLAN Y Evaluación De Su Capacidad.....	45
4.2.1 Montaje del Sistema.....	46
4.2.2 Métodos de Evaluación.....	49
4.2.3 Despliegue de resultados.....	50
4.2.3.1 Resultados obtenidos del trabajo con ITG.....	50
4.2.3.2 Resultados obtenidos del trabajo con Asterisk.....	54
4.2.3.2.1 Resultados con G.711.....	55
4.2.3.2.2 Resultados con G.729.....	58
4.2.3.2.3 Resultados con G.723.....	59
4.2.3.3 Análisis de resultados.....	61
4.3 Lineamientos para la selección de Parámetros Para Comunicaciones De VoIP Sobre 802.11.....	63
4.3.1 Selección de los mejores parámetros en redes no idóneas para transmisión de VoIP.....	64
4.3.1.1 Enlaces de Ancho de Banda limitado.....	64
4.3.1.2 Enlaces con Altos Retardos.....	66
4.3.1.3 Enlaces con Pérdidas de paquetes constantes.....	66
4.3.2 Comprobación y validación en laboratorio.....	67
4.3.2.1 Escenario 1: Canal de 28.800 bps, pérdidas del 1%, retardo de 100ms (enlace conmutado V.34).....	68
4.3.2.2 Escenario 2: Canal de 128.000 bps, pérdidas del 1%, retardo de 300ms (enlace de datos con múltiples saltos).....	69
4.3.2.3 Escenario 3: Canal de 64.000 bps, pérdidas del 5%, retardo de	

100ms (enlace de datos con múltiples saltos).....	70
4.3.3 Análisis de Resultados.....	70
4.4 Construcción Y Desarrollo De Un Sistema De Reconocimiento De Voz En Español.....	71
4.4.1 Requerimientos.....	71
4.4.2 Implementación.....	72
4.4.2.1 Preparación del reconocedor.....	73
4.4.2.1.1 Creación de la gramática.....	73
4.4.2.1.2 Creación de la red de palabras y el diccionario a partir de la gramática.....	74
4.4.2.1.3 Creación de Monófonos.....	74
4.4.2.1.4 Obtención de las grabaciones de entrenamiento	75
4.4.2.1.5 Generación de archivos de transcripción.....	75
4.4.2.1.6 Parametrización de las señales de audio	76
4.4.2.2 Creación de modelos ocultos de Markov.....	77
4.4.2.2.1 Generación de HMM.....	77
4.4.3 Evaluación de resultados.....	79
4.5 Selección de las mejores condiciones y parámetros para efectos de un óptimo reconocimiento de la señal de audio	80
4.5.1 Evaluación de reconocimiento de voz con bajas pérdidas con diferentes codificaciones.....	80
4.5.2 Evaluación de reconocimiento de voz con pérdidas con diferentes codificaciones.....	81
5 Esquema de implementación de Sistemas de Reconocimiento de Voz sobre VoWLAN para el idioma español.....	83
5.1 Propuesta de Esquema de operación para solución de reconocimiento de voz operando sobre VoWLAN.....	83
5.1.1 Módulo Registro.....	84
5.1.1.1 Validación del cliente.....	84
5.1.1.2 Información de estado.....	84
5.1.1.3 Almacenamiento de información de registro incluyendo:.....	84
5.1.2 Módulo Control.....	85
5.1.2.1 Submódulo de identificación de parámetros de conexión.....	85
5.1.2.2 Submódulo de Control de Capacidad.....	85
5.1.2.3 Base de datos de estados de conexión.....	86
5.1.3 Módulo de Aplicaciones.....	86
5.1.4 Módulo de Servicios de Red.....	86
5.1.4.1 DHCP.....	86
5.1.4.2 AAA.....	87
5.1.4.2.1 Acceso a la red inalámbrica.....	87
5.1.4.2.2 Autenticación en el servidor de registro.....	87
5.1.4.3 DNS.....	87
5.1.5 Red de Acceso.....	87
5.1.5.1 Cobertura.....	88
5.1.5.2 Capacidad.....	88

5.1.5.3 Calidad del Servicio.....	89
5.1.5.3.1 Modo de Operación.....	89
5.1.5.3.2 Interferencias.....	89
5.1.5.3.3 Mínimos condiciones de asociación.....	89
5.1.6 Terminales.....	90
5.2 Prototipo de un sistema de reconocimiento de voz sobre VoWLAN.....	90
5.2.1 SIP Registrar.....	91
5.2.1.1 Módulo de registro simple.....	92
5.2.1.2 Módulo de medición de parámetros de calidad.....	92
5.2.1.2.1 Base de Datos de parámetros de calidad.....	93
5.2.1.3 Módulo de control de capacidad.....	93
5.2.2 Servidor de Aplicaciones.....	94
5.2.2.1 Entrenar:.....	94
5.2.2.2 Reconocer:.....	94
5.2.2.3 Llamar:.....	95
5.2.2.4 Consultar:.....	95
5.2.2.5 Sintetizar:.....	95
5.2.3 ASR.....	95
5.2.4 TTS.....	96
5.2.5 Acceso inalámbrico.....	96
5.3 Pruebas al prototipo.....	97
5.3.1 Descripción del sistema típico.....	97
5.3.2 Operación en condiciones promedio pérdidas muy bajas.....	99
5.3.3 Operación en condiciones promedio pérdidas muy bajas modo G....	100
5.3.4 Operación en condiciones promedio, con pérdidas debido a interferencias de canales adyacentes.....	101
5.3.5 Conclusión a resultados de las pruebas.....	102
CONCLUSIONES y futuras líneas de investigacion.....	103
Referencias.....	105
anexo a – configuración de asterisk, d-itg y generación automática de llamadas	108
Evaluación de la Capacidad con ITG.....	108
Evaluación de la Capacidad con Asterisk.....	110
Configuración del Servidor.....	111
Sección de Clientes SIP “sip.conf”.....	111
Sección de Plan de Marcado “extensions.conf”.....	111
Sección de Clientes IAX “iax.conf”.....	112
Configuración del Cliente.....	113
Sección IAX iax.conf.....	113
Generación de llamadas.....	113

RESUMEN

El uso de tecnologías de VoIP sobre redes inalámbricas wifi (VoWLAN) combina beneficios de las dos tecnologías y presenta una atractiva propuesta tanto para proveedores de servicio como para los usuarios del sistema. A pesar de esto, la puesta en operación de servicios de voz sobre ip, requiere de ciertas condiciones en la red donde se transporta que permitan ofrecer un servicio de adecuada calidad. Las redes wifi presentan altas probabilidades de pérdida de paquetes y por la naturaleza del medio de aire, pueden generar efectos como retardos y jitter que son de poca importancia para el tráfico de datos pero críticos en servicios de tiempo real como la voz, más aun cuando se requiere el procesamiento de la señal obtenida al otro extremo para efectos de reconocimiento del habla.

Mediante el presente proyecto, desarrollado con un enfoque experimental, se evaluó el desempeño de un sistema VoWLAN, identificando los parámetros y las condiciones necesarias para permitir una comunicación confiable y con una buena calidad en la señal de voz que permitiera realizar procesos de reconocimiento automático del habla sobre la misma. A partir de los hallazgos realizados a través de varios laboratorios y pruebas experimentales, se propone un esquema de implementación que permita llevar a cabo de forma exitosa dicha tarea. Se concluye con el desarrollo e implementación de un prototipo creado a partir del esquema propuesto y la validación del mismo a través de la exposición de resultados concluyentes, realizando así, varios aportes originales en cuanto a la identificación de condiciones y de parámetros que determinan la capacidad y calidad de señal transmitida en un sistema VoWLAN.

1 INTRODUCCION

1.1 JUSTIFICACION Y ANTECEDENTES

El uso de los sistemas de reconocimiento de voz y en general de las interfaces orales para la interacción humano-computador y humano-máquina son una excelente alternativa al uso de las interfaces tradicionales, pues proveen mecanismos de interacción naturales, intuitivos y al alcance de todo tipo de personas al usar el lenguaje natural como medio de comunicación. Estos sistemas brindan ventajas al eliminar la necesidad de una disposición exclusiva de la persona, permitiendo que ésta pueda estar realizando otras actividades mientras interactúa con la máquina e incluso, permitiendo la operación de la máquina a personas con limitaciones visuales o motrices. Estas interfaces, a pesar de prometer ser una excelente alternativa a las interfaces de usuario gráficas, están aun en etapa temprana de desarrollo, presentan una alta complejidad y cuentan con un limitado grado de confiabilidad [1][2].

En el momento se cuenta con diversos mecanismos y métodos para el reconocimiento de voz, que datan desde principios de los 80's hasta los más modernos apenas en proceso de investigación. La problemática que abarca la construcción de aplicaciones de este tipo radica en la complejidad, exactitud y velocidad de cada una de las técnicas, en las que actualmente se encuentran tasas de efectividad de reconocimiento cercanas al 99% para un conjunto limitado de palabras cuando se usa un ambiente sin ruido y con restricciones gramaticales para el locutor (algunos cientos de palabras) y cercanas al 96% para conjuntos muy grandes de palabras (alrededor de 50000) [2][3]siempre y cuando el habla

sea de palabra aislada o de palabra conectada, sin usar habla continua en un reconocedor con independencia del locutor. Actualmente se vienen realizando diversos trabajos en torno al mejoramiento de la tasa de reconocimiento en diferentes ambientes atacando problemáticas como ruido ambiental[4], ruido de codificación [5], ruido por pérdidas de muestras [6][7], lejanía de la voz[8], entre otros.

Es importante considerar otros factores diferentes a la tasa de reconocimiento, tales como el tiempo de procesamiento y la predicción de errores, pues en determinados ambientes, la respuesta en tiempo real puede ser necesaria y los mecanismos que proveen una mayor tasa de reconocimiento pueden tardar mucho tiempo en realizar un proceso confiable [9].

La operación de sistemas de reconocimiento del habla se observa ya en diferentes tipos de software y hardware como la marcación automática en teléfonos celulares o en sistemas domóticos para la operación por voz en el hogar.

El esquema típico de un sistema de interacción por medio de interfaces orales que usa sistemas de reconocimiento y síntesis de voz se puede observar en la figura 1.1[11]

De otro lado, existen dos tecnologías que han aumentado su popularidad desde hace algunos años, y las proyecciones indican que continuarán siendo tecnologías de rápida apropiación en el futuro, estas tecnologías son las Redes Inalámbricas (WLAN) y la transmisión de Voz por redes de Datos IP (VoIP), ambas tecnologías presentan altos beneficios y usadas en conjunto proveen grandes ventajas como movilidad, bajo costo, facilidad de uso y facilidad de desarrollo de nuevos servicios, pero a la vez suman sus dificultades y problemas como el manejo de la calidad de servicio, interferencias y cortes cuando se hace *roaming* entre puntos de acceso.

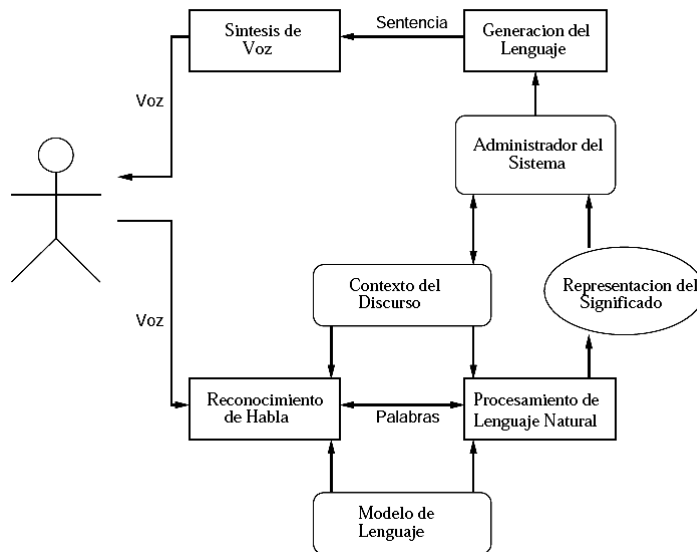


Figura 1.1 Diagrama en bloques de un sistema de interacción por medio de voz (Interfaz Oral) [9]

La implantación de sistemas de reconocimiento del habla en ambientes VoWLAN (VoIP sobre redes WLAN), abre la posibilidad de nuevos servicios y aplicaciones tales como sistemas de interacción a manos libres en campos abiertos y sistemas domóticos, los cuales hacen prever un amplio campo de investigación en el futuro inmediato.

Es así como se plantea la necesidad de una solución al problema del reconocimiento del habla, con la cual se garantice un alto grado de efectividad al usar el sistema en un ambiente VoWLAN, usando idioma español.

Debido a que el desarrollo de este proyecto comprende varias áreas de trabajo, se realizará énfasis en la búsqueda de la mejor solución para la minimización del impacto de la codificación, pérdidas y retardo en ambientes VoWLAN y en la tasa de reconocimiento de voz[12].

1.2 METODOLOGÍA

Para el desarrollo del presente proyecto se adoptó como metodología el método en cascada [13][14], sin embargo, para la parte de experimentación y pruebas, se adoptó el método científico, pues se adapta más a la naturaleza de las mismas.

Para la experimentación y desarrollo en el laboratorio se viene trabajando dentro de un ambiente de software libre donde se dispone de equipos de acceso inalámbricos como Access Points y Tarjetas de Red, además de teléfonos IP y Softphones. Para el desarrollo de las aplicaciones de reconocimiento de voz se usaron diversas herramientas, destacando HTK [15] con el cual se construyó el sistema de reconocimiento de voz, gracias a la facilidad que ofrece el mismo para crear y manipular modelos ocultos de Markov.

Para la experimentación se trabajó haciendo uso del método científico [16]

- Observación
- Inducción
- Hipótesis
- Prueba de la hipótesis por experimentación
- Demostración o refutación de la hipótesis
- Conclusiones.

1.3 ALCANCE Y LIMITACIONES

Para la implementación del prototipo se requirió de varios equipos y ambientes, en el caso de los equipos, se dispuso de la gran mayoría de ellos, ante la carencia de algunos, se suplió la funcionalidad requerida mediante software de comunicaciones, generadores de tráfico y otros medios. Los ambientes requeridos para mostrar la operación del prototipo son muy heterogéneos y se requeriría de muchos equipos de comunicaciones para la implementación. Por esta razón, se recrearan los ambientes de la forma más exacta posible, generando los fenómenos y condiciones típicas de cada uno de estos medios.

Debido a que el reconocedor construido, requiere de un entrenamiento previo a su

funcionamiento, se debió limitar la cantidad de palabras, delimitando el conjunto de vocablos ya que para aumentar su eficiencia es necesario ser entrenado por diferentes voces lo que tomaría excesivo tiempo y esfuerzo. Sin embargo para efectos de validación del esquema, se considera valida la delimitación del conjunto de vocablos.

Para la implementación del servicio de VoIP sobre WLAN, se limitó a 802.11b y 802.11g pues no se dispuso de equipos para trabajar con 802.11a

1.4 ORGANIZACIÓN

En el primer capítulo se realiza una introducción y descripción general del trabajo realizado.

En el segundo capítulo se encuentra una descripción y base teórica referente a las tecnologías de voz sobre IP y en general los aspectos fundamentales que atañen el uso de VoIP sobre redes WLAN, también llamado VoWLAN.

En el capítulo tres, se explican los mecanismos existentes para realizar reconocimiento de patrones y en especial los más adecuados para el reconocimiento de la voz.

En el cuarto capítulo, se muestran los principales trabajos, simulaciones y laboratorios realizados a partir de los cuales se obtienen las principales consideraciones para obtener una buena calidad de la señal de audio transmitida y mitigar en lo posible efectos nocivos como pérdida de paquetes y deterioro de la señal por codificación, demostrando como estas condiciones redundan en un mejor desempeño del sistema de reconocimiento.

En el capítulo cinco se propone un esquema completo para optimizar el reconocimiento de voz en un sistema de VoWLAN y se finaliza con la descripción del montaje del prototipo implementado sobre el cual se realizan diversas pruebas que validan la efectividad del modelo.

2 VOWLAN

En el presente capítulo se describen aspectos referentes al funcionamiento de tecnologías de VoIP y las principales incidencias de estos al operar sobre sistemas de VoWLAN. Con esto se busca introducir al lector a las tecnologías usadas como centro de evaluación del proyecto.

2.1 VOIP

VoIP (Voz sobre Protocolo de Internet) es la tecnología que permite digitalizar la señal de voz, codificarla, enviarla a través de una red de datos con protocolo IP y en la recepción decodificarla y convertirla de nuevo en señal analógica de voz. Para que tal proceso sea posible, se requiere el trabajo a diferentes niveles y con diferentes protocolos que permitan desde la digitalización de la señal, la señalización del proceso de llamada, la paquetización de las muestras y el transporte de la señal entre otros.

2.1.1 Protocolos VoIP

Los protocolos usados en VoIP se pueden dividir en tres grupos que serán explicados a continuación [17][18]:

- Protocolos de Señalización
- Protocolos de Transporte
- Protocolos de Enrutamiento

2.1.1.1 Protocolos de Señalización

Para realizar la llamada de un extremo a otro, es necesario el uso de protocolos de señalización que entre otras funciones, son los responsables de registrarse en el servidor IP, enviar señales como inicio, ack, final, invitación, negociar contenidos o tipos de datos y enviar direcciones origen y destino, entre otros [17]. Actualmente dos protocolos son los más usados H.323 y SIP

2.1.1.1.1 H.323

La ITU-T desarrolló el primer estándar para el envío de contenido multimedia sobre redes de datos llamado H.323, que incluía el transporte de video, voz y datos. H.323 no fue diseñado específicamente para VoIP, a pesar de esto, actualmente, y en gran parte debido a la estandarización del protocolo, el mayor número de redes de VoIP desplegadas en el mundo usan este protocolo de señalización.

H.323 está conformado de diferentes subprotocolos, H.225.0 para el registro, admisión, control y señalización de la llamada, H.245 para la descripción del tipo de servicio, intercambio de las capacidades de los terminales y el control lógico del canal que lleva los flujos multimedia.

H.323 también realiza el control de la llamada negociando codecs de audio y video y usa RTP para el envío de datos multimedia en tiempo real como la voz y el video.

2.1.1.1.2 SIP

Así como H.323, El protocolo de Inicio de Sesión, SIP, realiza el establecimiento, modificación y terminación de sesiones multimedia. Maneja servicios en tiempo real incluyendo la voz, el video, y otros como juegos y mensajería instantánea. El protocolo fue diseñado para ser muy flexible y se vale de otros protocolos para su funcionamiento en un ambiente de producción, tales como directorios LDAP, servidores AAA Radius y RSVP para proveer calidad de servicio, entre otros. SIP generalmente se acompaña de RTP

para realizar el transporte de la voz, sin embargo puede usar otro protocolo para tal fin, e interactúa directamente con el protocolo de descripción de sesión SDP y el protocolo de anuncio de sesión SAP. Debido a que SIP fue diseñado para operar en redes IP, fue desarrollado para manejar una sintaxis similar a http y smtp, compartiendo incluso, varios códigos de estado de http.

2.1.1.2 Protocolos de Transporte

2.1.1.2.1 RTP

El Protocolo en tiempo real, RTP, es el protocolo más usado para comunicaciones de voz, trabaja generalmente usando UDP para el transporte, por lo cual es sensible a pérdidas al no realizar retransmisión además de no garantizar el arribo de paquetes en el orden de despacho. Para sortear este problema, los paquetes RTP incluyen una marca de tiempo **TIMESTAMP** que indica el momento en que fue muestreado el paquete en el origen. En el destino se usa esta marca para sincronizar los paquetes en orden de despacho y determinar el retardo y *jitter* de la comunicación. RTP es usado tanto en sistemas multicast como unicast y trabaja de la mano de protocolos como SIP y H.323 en sistemas de VoIP.

El uso de RTP esta acompañado de RTCP. Cada vez que se establece una comunicación con RTP, también se establece una usando RTCP. La idea de RTCP es la de dar una realimentación sobre la calidad de la comunicación realizada, proveyendo información útil sobre el estado de la sesión RTP y facilitando la corrección de problemas en la comunicación.

2.1.1.2.2 CRTP

El proceso de paquetización de datos multimedia como la voz, requiere la inclusión de estos como carga útil de los paquetes. En el caso del audio, el tamaño de la carga útil por paquete suele ser pequeño y por lo tanto poco eficiente pues el tamaño de las cabeceras suelen ser iguales e incluso mayores. Una cabecera IP tiene de tamaño 20 Bytes, a eso se le suman 8 y 12 bytes mas de UDP y RTP respectivamente para una cabecera total de 40 bytes. Típicamente la cantidad de datos en una comunicación de voz usando g.729

introduce como carga útil 20bytes por paquete. Debido a esto, existen mecanismos de compresión de las cabeceras IP/UDP/RTP que permite que se compriman estos 40bytes entre 2 y 5 bytes, aumentando sustancialmente la eficiencia. CRTP implementa esta compresión, sin embargo, para que sea posible su uso, se requiere que todos los equipos por los que se enruten los paquetes soporten este protocolo.

2.1.1.3 Protocolos de Enrutamiento

El tráfico de VoIP es enrutado usando diferentes protocolos. Cuando se maneja tráfico en tiempo real como la voz, es necesario el establecimiento de rutas poco costosas, que proporcionen las mejores condiciones para el tráfico en tiempo real. A continuación se explican los protocolos mas usados, entre ellos, métodos de asignación de ruta estáticos y dinámicos y protocolos orientados a calidad de servicio [19].

2.1.1.3.1 RIP

Protocolo de enrutamiento de información, obtiene su métrica y decisión de ruta basándose exclusivamente en el número de saltos hasta el destino sin tener en cuenta parámetros como ancho de banda o retardos. La versión actual es la 2 y a pesar de sus limitaciones, este protocolo sigue siendo usado. El cálculo del costo de la ruta puede no ser el más adecuado al solo tener en cuenta el número de salto.

2.1.1.3.2 OSPF

Protocolo abierto de enrutamiento para redes ip basado en el algoritmo SPF, surgió debido a la incapacidad de RIP de manejar redes grandes y heterogéneas. Funciona enviando LSAs (Link Statement Advertisements) a los demás enrutadores en su misma área, los cuales incluyen información del estado del enlace.

2.1.1.3.2.1 SPF

Es el algoritmo de enrutamiento que usa OSPF. El enrutador envía a todos sus vecinos el estado de sus enlaces a través de sus interfaces y a su vez, recibe las bases de datos de sus vecinos, de modo que detecta fácilmente las fallas o caídas de rutas y establece unas nuevas. Cada uno de los enrutadores envía keep-alives continuamente, informando su estado operacional y reportando a sus vecinos cualquier cambio en él o en sus vecinos.

2.1.1.3.3 BGP

Border Gateway Protocol BGP permite el enrutamiento entre diferentes dominios en una red TCP/IP. BGP es un protocolo de gateway exterior usado comunmente en grandes ISP que tienen sus sistemas autónomos internamente y usan BGP para publicar sus rutas. A diferencia de los otros protocolos, BGP no usa el número de saltos o ancho de banda para escoger la ruta, sino una base de políticas de red determinadas previamente por el administrador.

BGP mantiene tablas de enrutamiento y transmite actualizaciones. La función principal de un sistema BGP es intercambiar la información de estado de los enlaces, incluyendo las rutas a los sistemas autónomos y compartir esta información a otros sistemas BGP. Con esta información se construye un esquema de conectividad de sistemas autónomos con los cuales se pueden tomar decisiones en las rutas a alcanzar.

De esta forma, cada router BGP mantiene una tabla de enrutamiento que muestra todas las posibles rutas hacia cada red en particular. Cuando un nuevo enrutador es conectado a la red, recibe las tablas de enrutamiento BGP completas, de allí en adelante, todos reciben actualizaciones netamente incrementales, las cuales no se realizan en tiempos programados, sino cuando ocurre cambio en la tabla de alguno de los enrutadores vecinos.

2.1.1.3.4 RSVP

Resource Reservation Protocol, Se debe aclarar que no es un protocolo de enrutamiento, pero se integra a ellos cuando se requiere QoS. Este protocolo provee de mecanismos de calidad de servicio a una red para que determinados flujos de datos sean tratados de forma especial. A través de la red se comunican los requerimientos para el flujo y se clasifica el tráfico en uno de tres grupos:

- **Mejor Esfuerzo:** Este es el tipo de tráfico más común, como transferencia de archivos, correo electrónico, http, que no tiene requerimientos especiales y es tolerante a retardos.
- **Sensible a la tasa:** Tráfico que requiere una tasa sostenida de datos, también llamado servicio de tasa de transferencia garantizada, comunmente usado en servicios que fueron originarios de redes de conmutación de circuitos, como servicios de voz y que al ser trasladados a conmutación de paquetes, en este caso

VoIP, requieren de un ancho de banda disponible garantizado.

- Sensible a retardo: Existe tráfico en el que más que garantizar una tasa de transferencia sostenida, se debe tener un retardo máximo garantizado, como el caso de videoconferencias con codificación variable, en las que, por ejemplo pueden requerirse 2MB para video con poco movimiento hasta 7MB para video con constante cambio. En cualquiera de los dos casos se requiere de un retardo bajo y constante.

2.1.2 Procesamiento y empaquetamiento de Voz sobre redes IP

2.1.2.1 Digitalización de la voz

Después de captar la voz del micrófono, ésta debe ser convertida como primer paso de analógica a digital para su posterior procesamiento y empaquetamiento, el proceso mediante el cual se realiza la digitalización es llamado PCM o Modulación por pulsos codificados.

2.1.2.1.1 PCM

La modulación por pulsos codificados ocurre en 4 etapas:

- Muestreo.
- Compansión.
- Cuantificación.
- Codificación.

2.1.2.1.1.1 Muestreo

Con el objetivo de obtener una representación digital de una señal analógica se realiza un muestreo usando dos veces la frecuencia de la señal. Como la señal a tratar es una señal de voz típica de un canal telefónico en la cual la máxima frecuencia es de 3,4Khz , que son aproximados a 4Khz, se realiza un muestreo de 8Khz o 125us, generando 8000 muestras de amplitud de pulso, que serán suficiente para obtener una buena representación de la señal original cuando sea reconstruida.

2.1.2.1.1.2 Compansión

El proceso de compansión se realiza a la señal ya muestreada y consiste en la compresión de las muestras digitales con un algoritmo específico, usualmente se comprimen las muestras de mayor amplitud. Con este proceso se logra una compresión típica de 35%, generando un error mínimo.

2.1.2.1.1.3 Cuantificación

En este proceso se hacen discretos los valores muestreados. Debido a que se dispone de un limitado número de bits para representar las muestras y que los valores de muestra son infinitos, se aproxima cada valor al valor discreto más cercano. Este proceso de aproximación genera ruido de cuantización que afecta la calidad de la señal, la cantidad de este ruido es inversamente proporcional al número de bits usados, de modo que se debe buscar un equilibrio entre la calidad de señal y ancho de banda usado.

2.1.2.1.1.4 Codificación

En este proceso a cada valor cuantificado se le asigna un código binario, este proceso es realizado por un códec, los parámetros que diferencian un códec de otro son:

- Número de Canales (Mono, Estéreo, etc.)
- Frecuencia de Muestreo
- Resolución (Número de bits)
- Tasa de Bits (Ancho de Banda)
- Pérdida (Asociada al error)

2.1.2.2 Códecs

Códec, Codificador/Decodificador o Compresor/Decompresor. Realiza la codificación de la señal en un extremo, preparándola para ser transmitida, y realiza la decodificación en el otro extremo para la recuperación de la misma. Aunque existen códecs con pérdida cero, la mayoría usado en sistemas multimedia como transmisión de voz y videos, requieren de optimización en el tamaño y por lo tanto el descarte de datos poco importantes, lo que genera pérdidas de la fidelidad, pero a su vez una gran compresión de los datos.

Los códecs de voz indican la forma en que será codificada la señal de voz y convertida en pequeños paquetes de datos por cada segmento de voz codificado. Existen diferentes formas de realizar este proceso las cuales difieren entre ellas en la cantidad de “compresión” que se puede lograr, y en los tamaños del segmento de voz que se muestrea, lo cual a su vez, determina la cantidad de retardo, la fidelidad de la señal y el tamaño de la señal codificada. A continuación, se muestran los códecs más usados en aplicaciones de audio [20].

G.711

G.711 es la técnica más conocida de codificación y como tal, es un estándar de la ITU-T. G.711 es usada a nivel mundial como la codificación de voz en redes telefónicas tradicionales en conmutación de circuitos. La tasa de muestreo es de 8000 Hz y una cuantificación uniforme de 8 bits, proporcionando un flujo de bits de 64kbps. Existen dos variaciones de G.711, la ley U y la Ley A. La Ley U es usada en Norte América y Japón y la ley A en el resto del mundo. G.711 ofrece una alta calidad en la voz y presenta una MOS (Mean Opinión Score) de 4.3. A pesar de la buena calidad de la voz y de la simplicidad de la codificación, este códec presenta una gran desventaja y es el alto consumo de ancho de banda de 64kbps. Esto sin tener en cuenta que cuando se realice la paquetización en una red IP, fácilmente sobrepasara los 80kbps, lo cual resulta en un alto costo en la transmisión para los proveedores de red. El retardo de este códec es el más pequeño y esta dado por $1/8000\text{Hz}=0,125\text{ms}$

G.728

Este códec basa su funcionamiento en filtros y un conjunto 1024 de vectores acústicos. Cada vector contiene un grupo de elementos los cuales representan varias características de la excitación de la señal. El códec envía la información correspondiente a los coeficientes extraídos, la ganancia y un apuntador al vector seleccionado. En el destino, el códec usa el mismo conjunto de vectores acústicos para regenerar la señal. Debido a la gran cantidad de vectores acústicos, son necesarios 10 bits para el apuntador, que al trabajar con 5 muestras a 8000Hz, producen una tasa de $8000 \times 10/5 = 16\text{kbps}$. Por la naturaleza de este códec, se considera un códec de muy bajo retardo ($5/8000\text{HZ}=0.625\text{ms}$) y presenta un MOS de 3.9

G.723.1

El códec G.723 busca solucionar el problema del alto consumo de ancho de banda de G.711 y G.728. Funciona a tasas de 6.3kbps y 5.3kbps. El codificador trabaja sobre bloques de 240 muestras tomadas en 30ms y dependiendo de la tasa de compresión, la muestra será de 20 o 24bytes. A pesar de la alta compresión lograda, el códec presenta un retardo de 30 ms por bloque mas 7.5ms de look-ahead que corresponde al tiempo que toma revisar el bloque para determinar como codificarlo. Este códec presenta un MOS de 3.8 y presenta una excelente tasa de compresión a costo de un retardo elevado de 37.5ms en el envío, que agregándole 30ms de decodificación, generan un retardo total de codificación/decodificación de 67,5ms

G.729

El códec G.729 busca mantener una alta tasa de compresión, con buena calidad de sonido y un retardo medio. Funciona a una tasa de muestreo de 8000hz, y trabaja sobre bloques de 80 muestras tomadas en 10ms. Este códec comprime a tasa de 8Kbps y genera un retardo de 10ms más 5ms de look-ahead, para un retardo total de codificación/decodificación de 25ms. G.729 presenta un MOS de 4.0 y presenta una muy buena relación ancho de banda /calidad de voz.

2.1.2.3 Esquemas de Paquetización

Una vez se tiene la voz codificada, ésta debe ser transmitida como datos. En el caso de estudio de este trabajo, la voz se transmite sobre una red IP, donde cada paquete debe ser enrutado para llegar a su destino. Esto redundante en información adicional que debe acompañar a la carga útil de cada paquete correspondiente a la voz codificada. Dependiendo de la carga útil a manejar, la información adicional, llamada cabecera IP puede ser incluso más grande que la información correspondiente a la señal de voz o carga útil, por ejemplo, una cabecera IP tiene de tamaño 20bytes, si se usa UDP y RTP son 8bytes y 12bytes más para una cabecera total de 40bytes. Esta cabecera será incluida en cada uno de los paquetes a enviar, de modo que si se realiza codificación con g.729 y se introduce como carga útil una sola muestra por paquete, se tendrán paquetes de 50bytes, conformados por 40bytes de cabeceras y 10bytes de audio codificado en g.729. Dependiendo del esquema de paquetización elegido, será el incremento de ancho de banda y de retardo.

Un esquema de paquetización difiere de otro en:

1. Tamaño de cabecera y uso o no de compresión, aunque una cabecera IP/UDP/RTP usará 40bytes, es posible usar un mecanismo de compresión que lo reduzca entre 2 y 5 bytes siempre y cuando sea soportado por todos los equipos que participan en la transmisión de extremo a extremo
2. Códec usado, además del ancho de banda usado, el códec determinará el retardo generado que se sumará al retardo en la transmisión. Si el retardo en la transmisión es muy grande, se tiende a usar códecs de bajo retardo.
3. Número de muestras incluidas en cada paquete, para disminuir ancho de banda es posible incluir varias muestras en un solo paquete, esto a expensas del aumento del retardo y de la sensibilidad a la pérdida de paquetes. Entre más muestras por paquete, mayor cantidad de información sin transmitir en una eventual pérdida del paquete.

2.1.3 Medición de la calidad de la voz

2.1.3.1 MOS

Las empresas de telecomunicaciones usan un sistema de medición de calidad subjetiva llamada MOS o Mean Opinion Score que mide la calidad de las comunicaciones telefónicas [18]. Estas técnicas de medición están definidas en ITU-T P.800 y están basadas en la opinión de personas quienes escuchan una muestra de tráfico de voz y califican la calidad de la voz escuchada. Los escuchantes son sometidos a varias preguntas como cantidad de ruido, eco, distorsión, pérdidas, retardos, entre otros. Después de responder las preguntas, las personas califican las muestras de uno a cinco, donde cinco significa excelente y uno significa malo. Una calificación general de 4 es considerada como muy buena, similar a la calidad de voz lograda en la PSTN.

2.1.3.2 PSQM

Otro método de probar la calidad de la voz es la PSQM o Perceptual Speech quality mesurement [18]. PSQM está basado en la recomendación de la ITU-T P.861, que

especifica un modelo para mapear las señales de audio a sus representaciones en el cerebro humano. La percepción de calidad de la voz depende de una mezcla de partes objetivas y subjetivas y varía entre los diferentes esquemas de codificación y las topologías usadas para el transporte. En PSQM se realiza un análisis objetivo de las señales de audio procesadas

comparando la señal original con la señal procesada.

2.1.4 Elementos básicos de un sistema VoIP

Existen gran cantidad de equipos y servidores que pueden operar en sistemas de VoIP y que difieren entre ellos dependiendo de factores como el protocolo de señalización usado. Para limitarse a una descripción general, se dividirá en dos grupos:

- Softswitch en el que estarán agrupados elementos de servicio y equipos de control
- Terminales

2.1.4.1 SoftSwitch

El softswitch coordina el control de las llamadas, señalización y procesos que hacen posible la comunicación a través de una o varias redes. Una vez una llamada es establecida, el control de conexión garantiza que la llamada se mantenga hasta que el iniciador o el receptor libere la conexión. Otras funciones del control de la llamada incluyen el reconocimiento de señales como ocupado o no disponible, reconocimiento de la iniciación y terminación de llamada, además de usar estos datos para el proceso de facturación entre otros.

Una función adicional e intrínseca del softswitch es la de separar los elementos red hardware de los elementos de control software.

2.1.4.2 Terminales

Existen terminales de muy diversos tipos, tanto para SIP como H.323, los más comunes son los teléfonos IP, los adaptadores telefónicos analógicos y los softphones

2.1.4.2.1 Teléfonos IP

Son teléfonos en apariencia, similares a teléfonos digitales encontrados en sistemas PBX

tradicionales, tienen interfaces de red ethernet o wifi y se registran contra servidores de VOIP normalmente SIP y H.323, los hay de diferentes tipos, desde teléfonos sencillos limitados a servicios de voz, hasta equipos con clientes de correo electrónico y capacidades de videoconferencia.

2.1.4.2.2 ATAs

Debido al gran número de sitios con cableado de cobre existente y a la necesidad de obtener una gateway económica y sencilla de VoIP a telefonía analógica, se crearon los ATAs. Los adaptadores telefónicos analógicos son equipos pequeños que realmente son clientes SIP o H.323 y se registran contra un servidor de VoIP y que presentan una interfaz con señalización FXS para conectar un teléfono tradicional de modo que el usuario final no notará que está usando VoIP. Estos equipos suelen ser muy económicos y presentan ventajas como la posibilidad de alojarse en un data center y estar protegidos con UPS o plantas eléctricas, ya que ellos alimentan la línea de cobre y el teléfono tradicional seguirá trabajando. Otro uso típico de estos equipos es en la conexión de equipos que requieren líneas de cobre tradicionales con señalización FXS como máquinas de FAX y datáfonos.

2.1.4.2.3 Softphones

Los Softphones son software que se instala en un computador, celular o dispositivo móvil el cual actúa como un cliente SIP o H.323, provee las mismas funcionalidades de un teléfono tradicional (algunas veces mayores) y sin necesidad de hardware adicional.

2.2 WLAN

A continuación se describirá el funcionamiento de las redes 802.11 también llamadas redes wifi o LAN inalámbricas. Se explicarán los elementos más importantes como la capa de control de acceso al medio MAC y la capa física. 802.11 define una MAC común y múltiples niveles físicos como 802.11a, 802.11b y 802.11g. Se limitará a las redes WLAN que operan en modo infraestructura pues las redes ad-hoc no son de interés en el presente trabajo y se finalizará con una explicación breve del funcionamiento de una

WLAN en modo infraestructura.

2.2.1 Acceso al medio

Antes de la transmisión de las tramas, se debe haber obtenido acceso al medio, en 802.11 hay definidas dos formas de acceso al medio [19]:

DCF o Función de coordinación distribuida (Para usarse en servicios con contienda)

PCF o Función de coordinación por punto (Para usarse en servicios libres de contienda)

A continuación se describirá el funcionamiento de DCF. PCF no está implementada y solo existe como especificación.

2.2.1.1 DCF

DCF se basa en CSMA/CA "Acceso Múltiple por detección de portadora con prevención de colisión". De esta forma, las estaciones wifi luchan por el acceso e intentan enviar tramas cuando detecten que no hay ninguna otra transmitiendo.

En detalle:

1. Antes de transmitir información a una estación se debe revisar el medio, para determinar su estado (libre / ocupado).
2. Si el medio no está ocupado, la estación ejecuta una acción adicional llamada espaciado entre tramas (IFS).
3. Si durante este intervalo temporal, el medio se determina ocupado, la estación debe esperar hasta el final de la transacción actual antes de realizar cualquier acción.
4. Una vez finalizada esta acción por encontrar el medio ocupado, la estación ejecuta un algoritmo de Backoff, que determina una espera adicional y aleatoria en un intervalo llamado ventana de contienda (CW). El algoritmo de Backoff entrega un número aleatorio y entero de ranuras temporales y su función es la de reducir la probabilidad de colisión que es máxima cuando varias estaciones están esperando

a que el medio quede libre para transmitir.

5. Mientras se ejecuta la espera marcada por el algoritmo de Backoff se continúa escuchando el medio de tal manera que si el medio se determina libre durante un tiempo de al menos IFS esta espera va avanzando temporalmente hasta que la estación consume todas las ranura temporales asignadas. En cambio, si el medio no permanece libre durante un tiempo igual o superior a IFS el algoritmo de Backoff queda suspendido hasta que se cumpla esta condición.

2.2.2 Nivel Físico

2.2.2.1 802.11b

802.11b surgió como una mejora a 802.11 que implementaba velocidades de 1 y 2 Mbps. 802,11b implementa velocidades adicionales de 5.5 y 11Mbps y trabaja sobre la banda de 2,4Ghz. Esta banda es usada por diferentes dispositivos y equipos lo cual representa una posible fuente interferente [19]. Otro problema reside en la existencia de solo 11 canales(en el estándar americano), el espectro en la frecuencia de 2,4 Ghz es de 84Mhz y un radio 802.11b usa cerca de 30Mhz cuando esta transmitiendo, de modo que para evitar interferencias entre puntos de acceso es necesario separar 3 canales, por ejemplo, usar los canales 1,6 y 11.

2.2.2.2 802.11g

802.11g surgió como una mejora a 802.11b que incluye velocidades de hasta 54Mbps en la banda de 2,4Ghz usando OFDM [19]. 802.11g es compatible hacia atrás con 802.11b. A pesar de ofrecer velocidades mucho mayores a las de 802.11b cuenta con las mismas limitaciones referentes al canal de radio pues opera en la misma banda y cuenta con el mismo número de canales. A pesar de que 802.11g es compatible con 802.11b, no es recomendable mezclar dispositivos pues al hacerlo, se tiene una degradación que disminuye la tasa de transferencia alrededor de 30% en todos los clientes que estén asociados al mismo punto de acceso. 802.11b y 802.11g usan diferentes tipos de modulación lo que no permite que interoperen directamente. Para permitir esta interoperatividad, es necesario incluir mecanismos de protección RTS/CTS request-to-

send/clear-to send. Estas trama RTS/CTS son incluidas en cada trama enviada, es por esto que se percibe una disminución en la tasa de transmisión de la red, este modo de operación es conocido comunmente como “modo mixto”. Para evitar este fenómeno se prefiere usar un modo “Solo G” y si es totalmente necesario admitir terminales 802.11b, se recomienda usar puntos de acceso separados 802.11g y 802.11b o puntos de acceso duales que trabajen por separado 802.11g y 802.11b. En estos dos casos, cada modo debe operar en canales distintos

2.2.3 Operación de la red en modo infraestructura

Una LAN inalámbrica en modo de infraestructura es la comunmente usada en sitios de acceso a internet inalámbricos, redes de empresa, hotspots, entre otros y busca extender una red cableada, permitiendo el ingreso de clientes inalámbricos. En este modo, los clientes inalámbricos se conectan a uno o varios puntos de acceso y estos a su vez a la red cableada.

Cada punto de acceso crea una celda de radio también llamada BSS o Basic Service SET en donde es posible la asociación de clientes inalámbricos que estén ubicados en la celda, el área de cobertura depende entre otros de la ubicación y tipo de antena, potencia de transmisión y capa física(802.11a/b/g). Otro factor a tener en cuenta es la tasa de datos deseada pues a mayor tasa, mayor relación señal a ruido es necesaria y por lo tanto, menor área de cubrimiento. Cuando no es necesaria una tasa mínima de transmisión, los clientes buscan negociar una tasa de transferencia acorde al SNR que presente en el momento. En el modo de infraestructura todo el tráfico enviado por un cliente inalámbrico, debe pasar por el punto de acceso al que está asociado, el cual se encarga de reenviar el tráfico a su destino, sea por medio inalámbrico o cableado. El estándar 802.11 permite que las celdas de radio estén superpuestas, translapadas o separadas, dependiendo de las necesidades de la red, se puede usar cualquiera de ellas, es importante aclarar que si se desean poner celdas de radio superpuestas, deben usar canales de radio espaciados con mínima interferencia entre ellos.

2.2.3.1 Escaneo

En una LAN inalámbrica, los clientes implementan una función de escaneo para poder encontrar los puntos de acceso en la zona de cobertura que se encuentra. Esta función es ejecutada al inicio de red del cliente y después de modo periódico para soportar el roaming. 802.11 define dos métodos de escaneo.

Escaneo Pasivo

Se sintoniza uno a uno cada uno de los canales, escucha cada uno por un periodo de tiempo y guarda información concerniente a cada punto de acceso encontrado. Este proceso se repite hasta haber abarcado todos los canales.

1. Por defecto cada punto de acceso transmite una trama de control cada 100ms con información del punto de acceso y estado.
2. El cliente que esté en ese momento sintonizando ese canal, recibe la información de control enviada por los puntos de acceso, determinando así que puntos de acceso están en su rango y en operación.
3. El cliente inalámbrico almacena el nivel de señal de los puntos de acceso y continua escaneando otros canales.
4. Después de escanear todos los canales el cliente decide a cual punto de acceso asociarse.

Normalmente el cliente decide asociarse al punto de acceso con mayor nivel de señal, sin embargo es posible incluir parámetros como niveles de ruido o cantidad de uso.

Escaneo Activo

1. El cliente envía señales de prueba en cada canal.
2. Los puntos de acceso que reciban la señal, responden enviando una señal de respuesta.
3. El cliente decide asociarse al punto de acceso con mayor nivel de señal y mejores condiciones

En algunos casos, algunos proveedores de equipos pueden implementar mejores o alternativas distintas para el escaneo activo, de modo que equipos diseñados para transmitir tráfico de voz, suelen tener mejoras para lograr una asociación y desasociación más rápida y así no tener un retardo muy alto en el caso de roaming.

2.2.3.2 Conexión a la red

Después de obtener una lista de puntos de acceso potenciales, el cliente procede a conectarse a la red sintonizando el canal del punto de acceso seleccionado e iniciar el proceso de asociación. Para iniciar la asociación, el cliente envía una trama de solicitud de autenticación, a la cual el punto de acceso responde. Este es el caso de autenticación abierta, también llamada *open system*. Es posible realizar autenticación con WEP o WPA, incluso usando 802.1X, para que esta sea realizada a través de servidores de autenticación. Después de realizar la autenticación, el cliente envía una solicitud al punto de acceso incluyendo datos como SSID, tasas de transferencia soportadas. Una vez el punto de acceso verifica el SSID devuelve una respuesta al cliente con un identificador de asociación y a partir de este momento se puede iniciar el tráfico de paquetes hacia otros equipos de la red.

2.2.3.3 Transferencia de Datos

La transferencia de datos entre el cliente y el punto de acceso es bidireccional.

A diferencia de una red Ethernet, es necesario un ACK por cada trama de datos que se envíe, esto debido a la naturaleza del medio de aire en donde es muy común la pérdida de paquetes. Si no se recibe un ACK después de un periodo específico de tiempo, la trama es retransmitida. Cuando muchas retransmisiones son requeridas, negocia una tasa de transmisión mas baja para hacer posible la comunicación.

2.2.3.4 Roaming

Cada cliente realiza un escaneo periódico, sea pasivo o activo, para mantener una lista actualizada de puntos de acceso. Sin en algún momento, la señal del punto de acceso al que esta asociado el cliente se torna muy débil, este inicia un proceso de reasociación con el nuevo punto de acceso y de deasociación del actual. El proceso de reasociación no requiere de autenticación, los datos referentes a esta son pasados directamente del punto de acceso anterior al nuevo.

2.3 VOWLAN

El estándar IEEE802.11 también conocido como wifi es muy similar al IEEE802.3 (ethernet), pues ambos están diseñados para operar en una LAN. Así como 802.3, 802.11 define un control de acceso al medio MAC común y varios niveles físicos como 802.11a, 802.11b y 802.11g.

Como tal, la operación de 802.11 es muy similar a 802.3 y se puede contar con los mismos elementos, solamente que en el caso de 802.11, se reemplazan los medios cableados por radios de dos vías de baja potencia. Limitaciones como el alcance y la velocidad alcanzada se definen por la potencia de los radios, ganancia de antenas, obstáculos en el medio, modulación usada, entre otros.

Partiendo de estos hechos, es posible realizar la transmisión de VoIP sobre 802.11, de forma similar a como se realiza con ethernet 802.3 pero con las limitantes de 802.11 y siguiendo una serie de recomendaciones para que las comunicaciones puedan realizarse de una forma confiable y con un mínimo de calidad establecido.

2.3.1 Consideraciones sobre WLAN

La interfaz de aire a través de la que se comunican los equipos wifi es bastante compleja y presenta bastantes dificultades para la realización de comunicaciones de voz, uno de los mayores problemas de VoWLAN es que el flujo de paquetes RTP/UDP puede ser perdido o interrumpido por diferentes fenómenos. Es por eso que al diseñar sistemas de VoWLAN se deben conocer las dificultades y limitaciones para poder minimizarlas y maximizar la calidad de las comunicaciones. A continuación se muestran las mayores problemas que atañen este tipo de sistemas:

2.3.1.1 Cobertura

La cobertura de la señal en una LAN inalámbrica define el espacio donde los usuarios

wireless pueden conectar a la red y usar efectivamente las aplicaciones que se prestan en ellas. Los puntos de acceso contribuyen a la cobertura formando celdas de radios que se traslapan entre ellas, las celdas normalmente tienen una forma irregular debido a los diferentes niveles de atenuación presentes en cada ambiente por donde se propaga.

Es importante definir las fronteras de cada celda, que dependen del ambiente donde se propagan, una vez se ubican eficientemente las fronteras, se puede buscar que las demás celdas del enlace wifi tengan el mínimo de traslape necesario para permitir roaming y en lo posible redundancia.

Se debe aclarar que la frontera está también definida por los requerimientos de la red. Por ejemplo, si es necesaria una conexión a una velocidad mínima de 11 Mbps, la frontera estará determinada en el punto donde la asociación aún alcance esta velocidad, a pesar de que en otro punto con menos nivel de señal se pueda asociar a 1Mbps

Para aplicaciones críticas o de considerable uso de ancho de banda, es importante establecer estas fronteras con altas tasas pues la tasa de transferencia real libre de interferencias importante no alcanza el 50% de la velocidad nominal, para una asociación a 54Mbps, la tasa máxima usualmente no supera los 24Mbps, con interferencias y pérdidas mínimas cercanas a cero

Cuando se usan aplicaciones de voz, la cobertura de una red inalámbrica debe definir el espacio donde los usuarios inalámbricos pueden conectarse a la red y usar las aplicaciones a un determinado mínimo nivel de calidad. La señal en el área de cobertura deberá permitir entonces que un teléfono IP inalámbrico se conecte a un punto de acceso y permita realizar una conversación con otro teléfono con una buena calidad del servicio y la comunicación. Para que esto sea posible, esto redundante en que se debe tener un mínimo de pérdidas y subsecuentes retransmisiones y que el canal tenga la capacidad disponible en el momento para permitir la comunicación. Es por esto que las fronteras no deben definirse teniendo en cuenta solo tasas de asociación, sino también ruido e interferencia existente, más aun en servicios de voz. Para tal efecto, se añade el factor

señal a ruido SNR como factor de decisión de fronteras. Este hecho permite que se garantice una buena calidad de servicio por lo menos desde el punto de vista de transmisión, pero, debido a la gran proliferación de redes wifi, el número de señales interferentes en una ciudad suele ser muy grande y mantener una SNR alta definitivamente acerca las fronteras y por lo tanto disminuye el tamaño de cada celda wifi. Según la literatura, un SNR capaz de proveer un óptimo rendimiento de una red inalámbrica debe estar entre 25 a 30dB.

2.3.1.2 Retardo

Debido al medio de acceso, las redes inalámbricas presentan retardos mucho mayores a los encontrados en sistemas que operan con medios cableados. Cuando se encuentran varios equipos asociados al mismo punto de acceso, deben tomar turnos para usar el medio y transmitir información. En una comunicación de voz, retardos grandes y variables podrían hacer la comunicación ininteligible e incluso perder la conexión. En el caso de roaming, el retardo será mayor mientras se desasocia de un punto de acceso y se asocia al otro, bajo adecuadas condiciones este proceso toma alrededor de 100ms, lo cual será notado en una comunicación de voz pero no afectará la inteligibilidad de la misma.

2.3.1.3 Interferencia

Existen diferentes fuentes de interferencia con las LAN inalámbricas que pueden afectar el rendimiento y operación. En casos de alta interferencia, la tasa de retransmisión puede ser superior al 20% e incluso al 50%, haciendo inviable el uso de aplicaciones de voz sobre la red [20]. Existen diversas fuentes de interferencias como hornos microondas, teléfonos inalámbricos y sistemas bluetooth que operan en la banda de los 2.4GHz, sin embargo el impacto de estos es limitado, pues el área de interferencia de los hornos no sobrepasa los 5mts, los sistemas bluetooth usan baja potencia y los teléfonos inalámbricos

Las mayores y más importantes fuentes de interferencia son presentadas por otros sistemas de LAN inalámbricos, por lo cual se debe hacer un análisis completo de interferencias en los sitios para ubicar los canales de operación de la mejor forma posible.

2.3.1.4 Seguridad

Aunque es posible encriptar la voz de forma previa a su transmisión, en un sistema VoWLAN se busca que el sistema de transmisión inalámbrica lleve a cabo esta labor y se deje como opcional una encriptación adicional a un nivel superior. De este modo, los mecanismos de seguridad se limitan a los mismos que provee el acceso a la WLAN, como WEP, WPA y 802.1X. Se recomienda usar un mecanismo de encriptación fuerte como 3DES o AES.

3 SISTEMAS DE RECONOCIMIENTO DE VOZ

3.1 FUNCIONAMIENTO DE UN SISTEMA DE RECONOCIMIENTO

La creación de un sistema de reconocimiento comprende las siguientes fases:

- Extracción de Características (Descriptores)
- Depuración de Características (Opcional)
- Clasificación

En la implementación de estos sistemas, normalmente se lleva a cabo primero un preprocesamiento de la señal donde se realizan filtrados y depuraciones de la muestra.

En el caso de aprendizajes supervisados, el proceso no termina en la clasificación, sino que se usa el resultado para procesos de entrenamiento a través de la verificación de mismo.

3.1.1 Extracción de Características (Descriptores)

El uso de un descriptor permite abstraer las características más importantes de una señal para efectos de una posterior clasificación. La selección de las características a extraer debe ser acorde al tipo de clasificación a realizar. A continuación se ilustrará el proceso de descripción usando el ejemplo de un OCR sencillo que se observa en la Figura 3.1

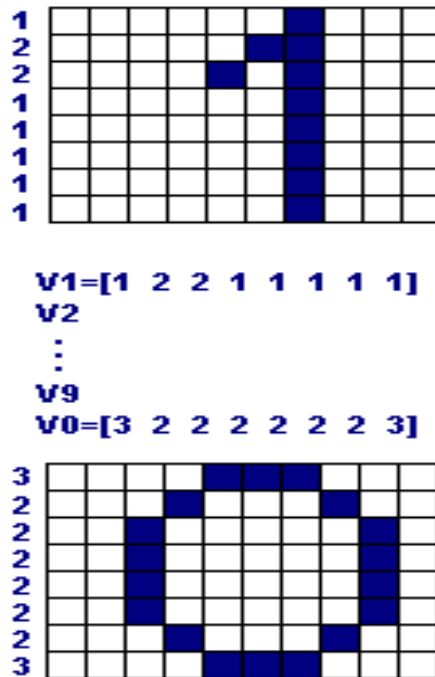


Figura 3.1 Extracción de vector de características

Como resultado del proceso de descripción, se deben obtener los vectores de características para cada elemento. En el ejemplo ilustrado en la figura 2, se muestra un descriptor creado para extraer características de dígitos. Se inscribe el dígito en una matriz y se cuenta el número de casillas pintadas por cada fila, esto genera un vector de n posiciones que corresponde al descriptor de la imagen.

3.1.1.1 Descriptores de Audio

El proceso de extracción de características es de gran importancia para obtener un buen resultado final. Es necesario seleccionar una serie de parámetros de la señal de audio que sean suficientemente representativos en aras de realizar una adecuada clasificación. Dicha selección de parámetros se realiza mediante los descriptores desarrollados específicamente para describir señales de audio.

A continuación se muestra el descriptor de audio mas usado en reconocimiento del habla.

3.1.1.1.1 Extracción de Características de Señales de Audio Mediante Coeficientes Cepstrales (MFCC Mel Frequency Cepstral Coefficients)

La señal de voz es muy variable en el tiempo y por lo tanto se debe descomponer en intervalos sobre los cuales se pueda definir estacionariedad. En el caso de la voz, estos intervalos de tiempo deben estar relacionados con la velocidad a la que el tracto vocal pueda modificar significativamente la señal. En aplicaciones prácticas se usan tiempos entre 10 y 20 msg.

Los coeficientes cepstrales de mel se usan para representar las señales de audio, entre ellas el habla, a partir de un modelo de bandas de frecuencia situadas logarítmica mente para asimilarse a la respuesta del oído humano.

El mecanismo para obtener la estructura armónica de la señal se conoce como liftering, el cual es útil, para el llamado aplanamiento espectral, en donde se remueve la envolvente y se pueden analizar las componentes armónicas de la señal[24].

Para extraer los coeficientes cepstrales en escala de mel o MFCC se siguen los siguientes pasos [25]:

1. Se toma la Transformada de una señal (una ventana)
2. Se mapea la energía del espectro obtenida en la escala de mel (logarítmica)
3. Se toma el logaritmo de la energía en todas las frecuencias de mel
4. Se toma la transformada coseno discreta de los logaritmos de energía en escala de mel como si fuera esta la señal.
5. Las amplitudes del espectro resultante serán los coeficientes.

Para efectos de reconocimiento usualmente se usan 13 coeficientes.

3.1.2 Clasificador

A partir de los vectores de características se clasifica la muestra en alguna de las clases

En la figura 3.2 se ilustra el funcionamiento de un clasificador de frutas, siguiendo dos descriptores.

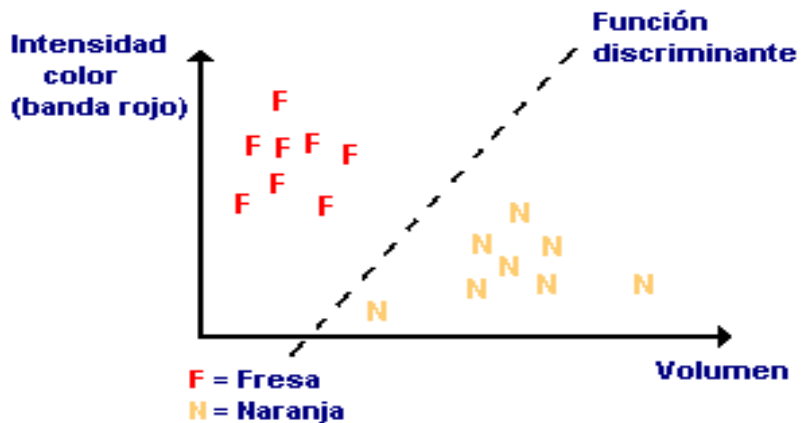


Figura 3.2 Ejemplo de un clasificador de frutas

Los dos descriptores usados en este ejemplo son la intensidad de color en la banda del rojo de la fruta y el volumen de la misma. El clasificador funciona a través de una función discriminante en la que los elementos de pequeño volumen y de color cercano a rojo serán clasificados como fresas y los de gran volumen y de baja intensidad en la banda del rojo serán clasificados como naranjas.

3.1.2.1 Tipos comunes de clasificadores

3.1.2.1.1 Clasificador por distancias euclideas

Este clasificador funciona calculando la distancia hacia cada clase desde la muestra a clasificar. La muestra es clasificada en el grupo que tenga una menor distancia hacia la muestra

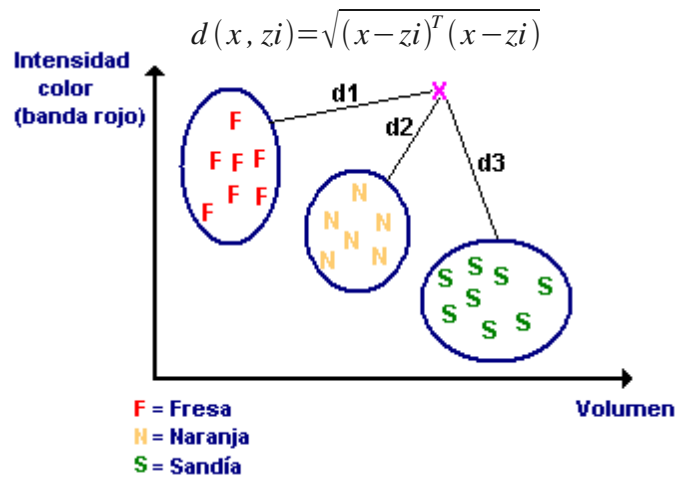


Figura 4. Ejemplo de un clasificador de frutas por distancias euclídeas

En la figura 4, se ilustra un clasificador de frutas. La muestra x es clasificada como naranja pues tiene la menor distancia euclídea hacia ese grupo de frutas.

3.1.2.1.2 Clasificador por vecino más cercano

Este clasificador funciona encontrando la menor distancia desde la muestra hacia cada una de los elementos de cada clase. La muestra se clasifica en el grupo donde este el elemento con menor distancia hacia la muestra

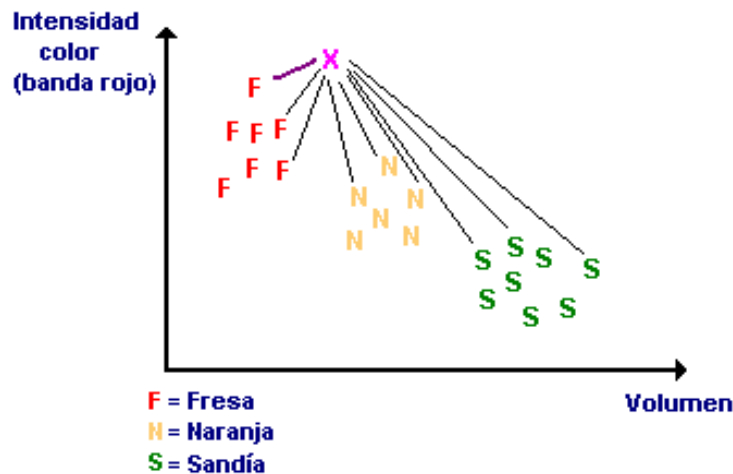


Figura 3.3 Ejemplo de un clasificador de frutas por vecinos cercanos

En la figura 3.3 se ilustra un clasificador por vecino más cercano. La muestra se clasifica

como fresa.

3.1.2.1.3 Clasificador por los k-vecinos más cercanos

Este clasificador funciona seleccionando los k elementos más cercanos a la muestra y la clasificación es determinada por la clase predominante

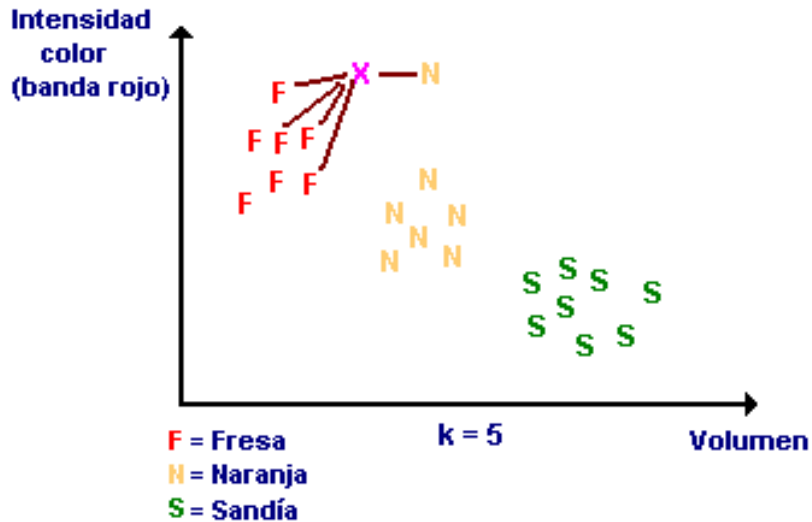


Figura 3.4. Ejemplo de un clasificador de frutas por k-vecinos más cercanos

En la figura 3.4 se ilustra un clasificador por k-vecinos más cercanos. La muestra se clasifica como fresa pues de las 5 distancias más pequeñas ($k=5$), cuatro corresponden a fresas y una a naranja

3.2 RECONOCIMIENTO DEL HABLA

El objetivo general de los sistemas de reconocimiento del habla es el de hacer cooperar un conjunto de informaciones en presencia de ambigüedades, incertidumbre y errores para obtener una interpretación del mensaje acústico recibido. La figura 3.5 muestra el funcionamiento modular de un sistema de reconocimiento del habla que a partir de diversas fuentes de conocimiento llega a una interpretación de la señal vocal introducida.

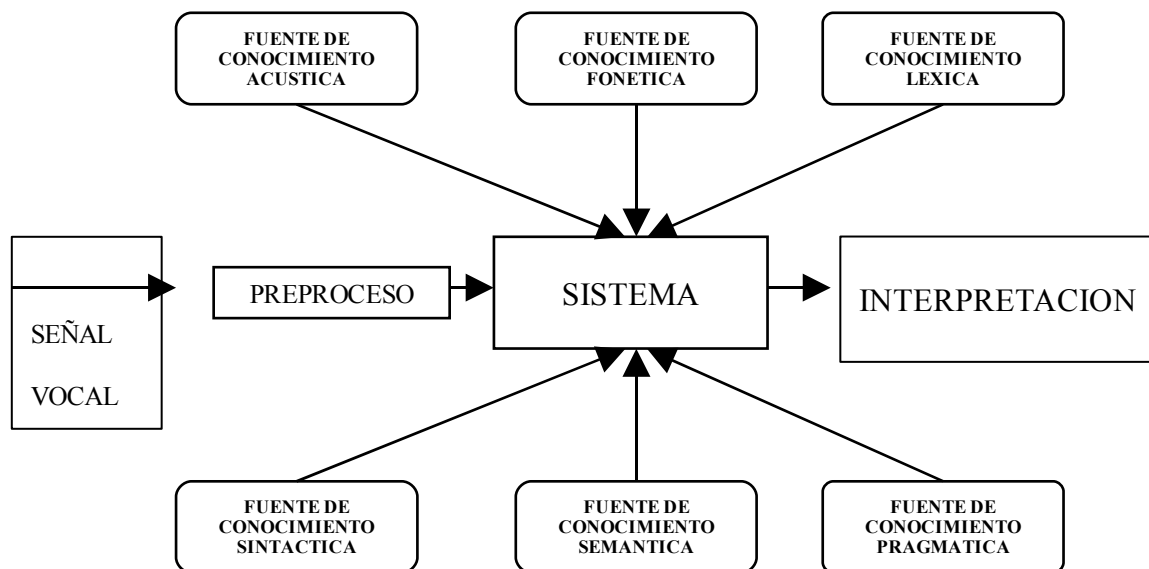


Figura 3.5. Sistema de Reconocimiento del habla RAH

3.2.1 Clasificación de sistemas de reconocimiento del habla

RECONOCIMIENTO DE PALABRAS AISLADAS.

Monolocutor y multilocutor (Dependiente e Independiente del Locutor). Léxico reducido (< 500 palabras).

RECONOCIMIENTO DE PALABRAS CONECTADAS.

Monolocutor y multilocutor (Dependiente e Independiente del Locutor). Sin sintaxis o muy simple. Léxico muy reducido.

REC. PAL. AISLADAS O CONECTADAS:

GRANDES DICCIONARIOS. Léxico grande (hasta 50.000 palabras)

RECONOCIMIENTO/COMPRESION DEL HABLA CONTINUA.

Léxico grande o medio. Tareas específicas. Sintaxis artificial o pseudo natural.

3.2.2 Técnicas de reconocimiento del habla

REGLAS (Sistemas Basados en el Conocimiento).

Aprendizaje deductivo

PLANTILLAS (Alineamiento Temporal No Lineal)

Aprendizaje por agrupamiento

PROBABILISTICOS (Modelos De Markov Ocultos)

Estimación de probabilidades

ESTRUCTURALES (Redes Finitas/Gramáticas)

Inferencia Gramatical y estimación de probabilidades

CONEXIONISTAS (Redes Neuronales Artificiales)

Estimación de pesos

3.2.3 Modelos Ocultos de Markov en el reconocimiento del habla

Los modelos ocultos de markov son modelos estocásticos que constan de un proceso oculto y un proceso observado cuyos estados dependen de los estados de los procesos ocultos.

La tarea fundamental consiste en determinar los parámetros ocultos a partir de los parámetros observados. La diferencia fundamental respecto a un modelo de Markov habitual consiste en que los estados no son directamente visibles para el observador, pero sí lo son las variables influenciadas por el estado. Cada estado tiene una distribución de probabilidad asociada sobre el conjunto de posibles valores de salida. La secuencia de valores de salida generados a partir de un HMM dará información sobre la secuencia de estados [26].

Para entenderlo mejor se ilustrará un proceso oculto de markov a través de un ejemplo:

Se tiene una mascota, de la cual no se puede conocer su estado de ánimo pues es un animal y no puede ser cuestionado sobre el mismo, sin embargo, se conocen factores generales sobre los que se podría llegar a deducir el estado de ánimo de la mascota. En este caso, el estado de ánimo representaría el modelo oculto. Ahora, el estado de ánimo es cambiante, y para efectos de facilidad del ejemplo, se va a representar como un modelo de dos estados, “alegre” y “triste” que representarían la cadena de markov, que está oculta pues no se puede ver directamente, de modo que se deben tomar algunas observaciones para poder deducir el estado de ánimo, en este caso, las observaciones serían “mueve la cola”, “se echa” y “ladra”.

Todos los elementos anteriormente mencionados conforman el modelo oculto de markov HMM y se resumirán a continuación:

- estados:
“triste”, “alegre”
- observaciones:
“mueve la cola” “se echa” “ladra”
- probabilidad inicial:
“triste 0,3” “alegre 0,7”
- probabilidad de transición:
De triste a (“triste 0,5” - “alegre 0,5”)
De alegre a (“triste 0,3” - “alegre 0,7”)
- probabilidad de emisión:
“Triste” (“mueve la cola 0,1” “se echa 0,6” “ladra 0,3”)
“Alegre” (“mueve la cola 0,7” “se echa 0,1” “ladra 0,2”)

A partir de los datos, la descripción del modelo realizado y las observaciones, es posible deducir probabilísticamente cual es el estado de ánimo de la mascota.

3.2.4 Modelos del lenguaje

Un modelo de lenguaje es un mecanismo para definir la estructura del lenguaje, es decir, para restringir adecuadamente las secuencias de unidades lingüísticas más probables.

En general son útiles en aplicaciones que exhiban una sintaxis y/o semántica compleja.

Un buen modelo del lenguaje solamente debería aceptar con alta probabilidad frases correctas y rechazar o asignar baja probabilidad a aquellas secuencias de palabras incorrectas [27].

Para el reconocimiento del habla, los modelos más usados son los basados en n-gramas.

3.2.4.1 Modelo n-gramas

La probabilidad de una palabra $P(W)$ se puede descomponer como:

$$P(W) = \prod_{i=1}^n P(w_i | w_1 \dots w_{i-1}) = P(w_1) \cdot P(w_2 | w_1) \cdot P(w_3 | w_1 w_2) \dots \\ P(w_n | w_1 \dots w_{n-1})$$

Con $N=2$ tendremos Bigramas

$$P(W) = P(w_1) \cdot P(w_2 | w_1) \cdot P(w_3 | w_2) \dots P(w_n | w_{n-1})$$

Con $N= 3$ tendremos Trigramas

$$P(W) = P(w_1) \cdot P(w_2 | w_1) \cdot P(w_3 | w_2 w_1) \dots P(w_n | w_{n-2} w_{n-1})$$

Para un mejor entendimiento, se ilustrará con un ejemplo:

Suponiendo que el resultado de una decodificación acústica asigna probabilidades similares a las siguientes frases:

the{big,pig}dog

Si $P(\text{pig} \mid \text{the}) = P(\text{big} \mid \text{the})$ entonces la elección de una u otra depende de la palabra dog.

$P(\text{the pig dog}) = P(\text{the}) \cdot P(\text{pig} \mid \text{the}) \cdot P(\text{dog} \mid \text{the pig})$

$P(\text{the big dog}) = P(\text{the}) \cdot P(\text{big} \mid \text{the}) \cdot P(\text{dog} \mid \text{the big})$

como $P(\text{dog} \mid \text{the big}) > P(\text{dog} \mid \text{the pig})$ el modelo ayuda a reconocer la frase correctamente.

El problema de estos modelos radica en la necesidad de un gran número de muestras de aprendizaje necesarias:

Ventajas:

- Fácil implementación.
- integración sencilla con el proceso acústico.
- Algoritmos eficientes de reconocimiento.
- Aprendizaje automático.

Desventajas:

- Son muy generales: Necesaria adaptación a aplicaciones.
- Falta de muestras de aprendizaje.
- Sólo captura relaciones a corta distancia.

4 MODELOS DE EXPERIMENTACIÓN Y SIMULACION

En el presente capítulo se detallan los principales trabajos realizados previamente a la construcción del esquema y prototipo completo, y que a través de sus resultados validarán lo propuesto en el esquema final. Cada uno de los trabajos se realizó teniendo en mente que sería parte del prototipo final y por lo tanto las condiciones y requerimientos de cada uno de ellos fueron planteados pensando en una integración futura de los sistemas. Elementos como código o descripción de procesos en la elaboración resultaron demasiado extensos y fueron remitidos a los anexos. A continuación se describe el contenido del capítulo.

La sección 4.1 incluye con un compendio de mecanismos para proveer calidad de servicio sobre 802.11, validados experimentalmente. Esta sección no incluye mecanismos de QoS como lo son marcado de paquetes, diferenciación de servicios, entre otros, pues estas funcionalidades solo son proveídas en ciertos equipos y por lo general tienden a ser más costosos y/o a ser mecanismos propietarios, lo cual los deja fuera del alcance del proyecto. En esta sección se muestran mecanismos que pueden ser implementados en cualquier tipo de equipos, incluso sin soporte de QoS, pero que puede mejorar notablemente la calidad de la comunicación.

En la sección 4.2 se presenta una descripción del trabajo y los resultados derivados de la construcción de un sistema de telefonía IP, que opera sobre WLAN, con el objetivo principal de obtener la capacidad del sistema para manejar canales de VoIP, evaluando el

efecto del delay, jitter y pérdida de paquetes en el sistema y su consecuente relación con las limitaciones en capacidad. Se encuentra que la capacidad es afectada por el tipo de acceso al medio que usa WLAN, lo que lleva a evaluar la capacidad usando parámetros adicionales como codificación. Los resultados de esta fase fueron publicados en la revista *Sistemas & Telemática*, indexada por Colciencias.

En la sección 4.3 se describen las adiciones al montaje ya realizado, con el objetivo de encontrar las condiciones óptimas para el transporte de VoIP en redes de distinto tipo, incluyendo entre ellas, WiFi. Gracias al montaje experimental, se pudo validar y obtener lineamientos en la selección de parámetros tales como esquemas de paquetización de la señal de audio, tipo de codificación, tamaño de muestreo, retardo, codificación, entre otros, en pro de una correcta recepción de la señal de audio haciendo énfasis en la calidad. Con este trabajo adicionalmente se mostró la incidencia de varios parámetros en la resistencia del sistema a la pérdida de paquetes, la calidad de la comunicación y el consumo de ancho de banda. Los resultados fueron publicados en la revista *Ingenium*, indexada por colciencias.

En la sección 4.4 se describe el desarrollo de un reconocedor de voz en español y los resultados obtenidos, de modo que fuera posteriormente usado en el prototipo final..

En la sección 4.5 se encuentra mediante la implementación del reconocedor en el sistema de VoWLAN. que parámetros son los más importantes se deben tener en cuenta para obtener una buena tasa de reconocimiento, consignando resultados de diversas pruebas en tablas y gráficas.

4.1 MECANISMOS PARA PROVEER CALIDAD DE SERVICIO SOBRE 802.11

“Calidad de Servicio se refiere a la capacidad de la red para proporcionar un mejor servicio a tráfico seleccionado de varias tecnologías de red. QoS procura que el tráfico de tiempo real de aplicaciones multimedia y de voz reciba la más alta prioridad, el mayor

ancho de banda y el menor retardo en comparación con el tráfico de datos considerado como de ‘al mejor esfuerzo’ ”. [28]

Las tecnologías de calidad de servicio proporcionan la base para el éxito de las aplicaciones multimedia y de voz y para el contexto de este trabajo especialmente en enlaces de bajas prestaciones o condiciones inadecuadas, como WiFi con pérdida de paquetes, enlaces de cobre de bajo ancho de banda, enlaces satelitales con alto retardo, entre otros. Esta calidad está determinada por 3 factores principales:

- Códecs de Voz
- Esquemas de Paquetización.
- *Jitter y Jitter Buffer*

4.1.1 Códecs de Voz

Los códecs de voz indican la forma en que será codificada la señal de voz y convertida en pequeños paquetes de datos por cada segmento de voz codificado. Existen diferentes formas de realizar este proceso las cuales difieren entre ellas en la cantidad de “compresión” que se puede lograr, y en los tamaños del segmento de voz que se muestrea, lo cual a su vez, determina la cantidad de retardo, la fidelidad de la señal y el tamaño de la señal codificada. En la tabla 4.1 se pueden observar las principales características. A continuación se detalla la incidencia en cada uno de estos parámetros.

Codec	Tamaño Muestra	Tiempo Muestra	Tamaño Payload
G.711 (64 Kbps)	80 Bytes	10 ms	160 Bytes
G.729 (8 Kbps)	10 Bytes	10 ms	20 Bytes
G.723.1 (6.3 Kbps)	24 Bytes	30 ms	24 Bytes
G.723.1 (5.3 Kbps)	20 Bytes	30 ms	20 Bytes
G.726 (32 Kbps)	20 Bytes	5 ms	80 Bytes
G.726 (24 Kbps)	15 Bytes	5 ms	60 Bytes
G.728 (16 Kbps)	10 Bytes	5 ms	60 Bytes

Tabla 4.1 Principales características códecs ITU

4.1.1.1 Retardo

Antes de realizar la codificación el códec debe tener una cantidad mínima de datos de la señal, por lo cual debe almacenar cada segmento de datos en un buffer antes de ser procesado. De esta forma, ya existe un retardo mínimo en el proceso de codificación y decodificación. Por ejemplo, el *buffer* para realizar la codificación usando g.729 debe ser de 15ms, 10ms del tiempo de muestra y 5ms debidos al retardo *look-ahead* que es el tiempo que toma el algoritmo en revisar el segmento y determinar cómo codificarlo. A este retardo de 15ms, se le debe sumar el retardo que toma la decodificación en el lado de la recepción, que en este caso será de 10ms, equivalentes al mismo tiempo de muestreo. De tal modo, el retardo originado por codificación en una conversación usando el códec g.729 será de 25ms. Se puede observar, que el retardo debido al proceso de codificación es muy pequeño comparado al de g.723 que sería de 67.5ms (7.5ms de *look-ahead*). Se debe tener en cuenta que este retardo no solamente afecta la percepción en la conversación, sino que a mayor retardo, mayor posibilidad de generación de eco y otros fenómenos de deterioro de la comunicación.

4.1.1.2 Calidad de la señal

Es de notar que para lograr una buena compresión de la señal de la voz, necesariamente se afectará la fidelidad de la señal y por lo tanto se percibirá disminución en la calidad del audio. Codificaciones de alta compresión como g.723 tienden a ser percibidas como de mediana calidad [29].

Existen diferentes técnicas que inciden en la percepción de calidad de la señal de voz en la comunicación, las principales son:

- *VAD (Voice Activity Detection)*.

Técnica usada para reducir el envío de datos cuando uno de los interlocutores no está transmitiendo voz. Para tal efecto el algoritmo se encarga de detectar la ausencia de voz. Esta técnica tiene dos inconvenientes:

- Sonidos de muy baja intensidad tienden a ser detectados como silencios y por lo tanto no son transmitidos.
- Cuando se detectan silencios, se envía una señal de intensidad cero, lo cual confunde al interlocutor por lo que tiende a pensar que la llamada se ha

desconectado.

- *Ruido Comfort.*

Se usa generalmente acompañado de algoritmos VAD. Se genera un ruido aleatorio suave en lugar de silencios cuando hay ausencia de voz.

- *Suavizado en la pérdida de paquetes.*

En la conmutación de paquetes, bajo determinadas condiciones, la pérdida de datos o el arribo tardío de paquetes que se puede asumir como pérdida es un fenómeno frecuente. Esta situación genera vacíos en la reconstrucción de la señal en el decodificador, lo cual pasa desapercibido si el paquete perdido tiene una carga útil pequeña. Sin embargo, ante la pérdida de paquetes con cargas útiles grandes, o pérdida de paquetes continuos con cargas útiles pequeñas, el vacío que se genera es lo suficientemente grande para que se afecte la inteligibilidad de la conversación. Cuando las pérdidas son pequeñas, es posible usar técnicas para “suavizar” la señal. Técnicas comunes son la repetición del último segmento de voz recibido y el relleno promedio en los vacíos.

4.1.1.3 Esquemas de Paquetización

En la conmutación de paquetes y en este caso en la transmisión de la voz sobre IP, cada paquete debe ser enrutado para llegar a su destino, esto redundante en información adicional que debe acompañar a la carga útil de cada paquete. Dependiendo de la carga útil a manejar, la información adicional, llamada cabecera IP puede ser incluso más grande que la información correspondiente a la señal de voz o carga útil, por ejemplo, una cabecera IP tiene de tamaño 20bytes, si se usa UDP y RTP son 8bytes y 12bytes más para una cabecera total de 40bytes. Si se realiza codificación con g.729 y se introduce como carga útil una sola muestra por paquete, se tendrán paquetes de 50bytes, conformados por 40bytes de cabeceras y 10bytes de audio codificado. Para mejorar esta relación, se cuenta con dos opciones: aumentar el tamaño de la carga útil por paquete y/o disminuir el tamaño de las cabeceras del paquete. Para lograr estos cometidos se puede realizar:

4.1.1.3.1 Transmisión de varias muestras por paquete.

Para mejorar la relación es posible enviar varias muestras por paquete, por ejemplo, enviar 4 en cada carga útil con lo que se tendrán paquetes de 80bytes, conformados por

40bytes de cabeceras y 40bytes de carga útil, sin embargo, esto genera dos efectos negativos. El primero, que se debe esperar a que las 4 muestras sean codificadas antes de ser enviadas, aumentado considerablemente el retardo, en el caso de g.729 se aumenta el retardo total de 25ms a 100ms. El segundo, en una ocasional pérdida de paquetes, se perderán 4 muestras de voz, afectando notablemente la inteligibilidad de la comunicación.

4.1.1.3.2 Compresión de Cabeceras:

Otra posibilidad es la de realizar compresión de las cabeceras. Existen diversos métodos, entre ellos ROHC, CRTP, IPHC, los cuales comprimen los 40bytes de IP/UDP/RTP entre 1byte – 5bytes dependiendo del método y tipo de información presente en la cabecera. Al comprimir la cabecera, se genera un retardo, sin embargo eligiendo adecuadamente el método este se hace muy pequeño y en la práctica puede tomarse como cero.

4.1.1.4 Jitter y Jitter Buffer

Se le llama *jitter* a la diferencia en el retardo de extremo a extremo entre varios paquetes. En una comunicación realizada usando conmutación de paquetes, suele suceder que los paquetes tomen diferentes caminos para llegar a su fin. Cuando paquetes que fueron enviados primero llegan después que los últimos enviados, y no alcanzan a llegar antes de que se complete la muestra para ser decodificada, el *jitter* es tan alto que ocurre descarte de paquetes y se generan huecos en la señal a decodificar, afectando sensiblemente la regeneración de la señal. Para mitigar el efecto del *jitter*, se suele usar un *buffer* previo a la decodificación de la señal. Es de tener en cuenta también, que a mayor tamaño del *buffer*, mayor el retardo adicional se generará por efectos del mismo.

4.1.1.5 Consideraciones de QoS

Uno de los roles de la calidad de servicio es mantener el retardo, el *jitter* y las pérdidas de paquetes para los tipos de tráfico seleccionados dentro de límites aceptables. Los requerimientos que se deben cumplir cuando se trabaja con voz sobre IP son:

- Retardo máximo en un sentido no mayor a 150 ms (de acuerdo a la recomendación ITU-T g.114). Sin embargo, es importante considerar que los usuarios normalmente notarán los retardos de la voz si estos en viaje redondo

sobrepasan los 250ms [20].

- Pérdida de paquetes mínima: VoIP no es tolerante a las pérdidas de paquetes, aún con un 1% de paquetes perdidos se puede degradar enormemente una comunicación de voz así se esté utilizando el códec g.711, en caso de utilizar códecs con mayor tasa de compresión la pérdida es prácticamente intolerable [29] [30].
- El *jitter* promedio no debe ser mayor que 30ms [6] aunque algunos autores hablan de hasta 50ms [31][32].

4.2 CONSTRUCCIÓN DE UN SISTEMA DE VOIP SOBRE WLAN Y EVALUACIÓN DE SU CAPACIDAD

A la fecha, existen gran cantidad de redes inalámbricas de área local implementadas a nivel mundial sobre los estándares 802.11a/b/g pero que no proveen ningún tipo de control o acciones para ofrecer QoS puesto que son previas a la generación del actual 802.11e, incluso aún poco comercializado. Sin embargo, en varios de estos lugares se ha visto la necesidad de realizar la transmisión de VoIP sobre estas redes y es muy importante conocer los resultados que se obtienen al evaluarlo de forma experimental teniendo en cuenta una gran cantidad de factores que no se pueden tener en escenarios de simulación. En esta sección, se presentan los resultados de evaluar de forma experimental la capacidad de 802.11 para soportar comunicaciones de VoIP; para tal efecto se realizó la verificación de dos formas, usando el generador de tráfico ITG y haciendo la generación de llamadas reales usando la PBX-VoIP Asterisk. Como resultado, se observará que lo obtenido por ambas vías fue en general coincidente y que en las redes inalámbricas sin manejo de QoS el principal limitante en la capacidad y en el número de llamadas concurrentes es la contienda por el acceso al medio y no tanto el tipo de códec usado y/o el ancho de banda que consume cada llamada. Al final de la sección se pueden observar los resultados resumidos en tablas, que muestran la máxima cantidad de llamadas sin degradación de la calidad para cada códec usado. Se concluye también que en una red 802.11 la capacidad de canales de VoIP no puede calcularse ni aproximarse con una simple división entre ancho de banda total y ancho de banda por

canal, sino que deben considerarse más factores, pues, lo que se observa inicialmente es una degradación cuasi-exponencial al aumentar el número de comunicaciones hasta cierto punto y luego se tiene una degradación abrupta que incluso hace caer las demás comunicaciones de voz que se estén realizando en ese momento.

4.2.1 Montaje del Sistema

Para el montaje del sistema, y ante la necesidad de aproximarse a un entorno de operación real, se realizó el montaje tal como se indica en la figura 4.1., en donde se contó con 6 clientes conectados de manera inalámbrica a un punto de acceso y éste a su vez conectado por su interfaz Ethernet a 100 Mbps con el servidor de VoIP. Se trabajó en un ambiente heterogéneo en cuanto a modelos y marcas de equipos que trabajan con 802.11, así mismo los computadores usados tuvieron diferentes características hardware y ubicación en el laboratorio. El sistema operativo usado fue GNU/Linux, en dos distribuciones: Ubuntu y RedHat. En la figuras 4.2 y 4.3 se muestran fotografías de la infraestructura utilizada.

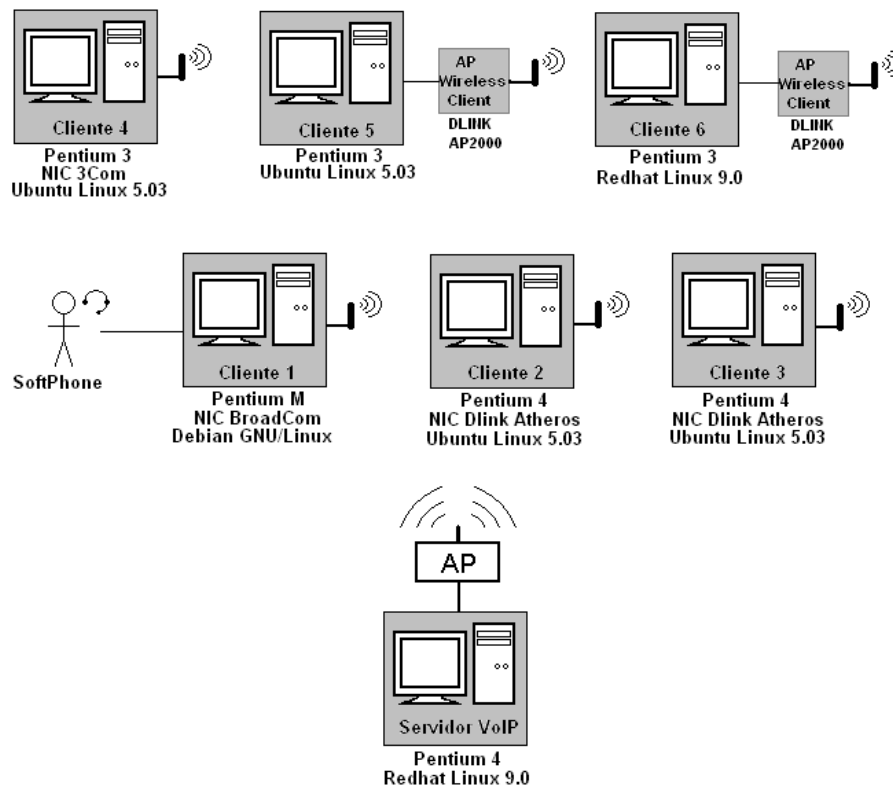


Figura 4.1. Laboratorio montado.



Figura 4.2 Infraestructura de Laboratorio Utilizada.



Figura 4.3 Equipo Cliente usando Softphone, Asterisk e Iptraf

La descripción de la infraestructura se resume en la tabla 2.

<i>Equipo/Características</i>	<i>CPU</i>	<i>Memoria</i>	<i>Red</i>	<i>Sistema Operativo</i>	<i>Software</i>	<i>Dirección IP</i>
Cliente1	Pentium M – 1.5Ghz	512MB	BroadCom Wireless	Debian GNU/Linux 3.1	Asterisk PBX 1.0.9 - ITG	192.168.1.100
Cliente2	Pentium 4 – 2.6Ghz	512MB	Dlink / Atheros Wireless	Ubuntu GNU/Linux 5.03	Asterisk PBX 1.0.6 - ITG	192.168.1.101

<i>Equipo/Características</i>	<i>CPU</i>	<i>Memoria</i>	<i>Red</i>	<i>Sistema Operativo</i>	<i>Software</i>	<i>Dirección IP</i>
Ciente3	Pentium 4 – 2.6Ghz	512MB	Dlink / Atheros Wireless	Ubuntu GNU/Linux 5.03	Asterisk PBX 1.0.6 - ITG	192.168.1.102
Ciente4	Pentium 3 – 800Mhz	128MB	3COM PCMCIA Wireless	Ubuntu GNU/Linux 5.03	Asterisk PBX 1.0.6 - ITG	192.168.1.103
Ciente5	Pentium 3 – 800Mhz	128MB	DLINK AP2100 Wireless Client Mode	Linux RedHat 9.0	Asterisk PBX 1.0.7 - ITG	192.168.1.104
Ciente6	Pentium 3 – 500Mhz	128MB	DLINK AP2100 Wireless Client Mode	Ubuntu GNU/Linux 5.03	Asterisk PBX 1.0.6 - ITG	192.168.1.105
Servidor VoIP	Pentium 4 – 2.6Ghz	512MB	Atheros Wireless	Ubuntu GNU/Linux 5.03	Asterisk PBX 1.0.6 - ITG	192.168.1.2
Punto de Acceso Linksys WRT54G	BCM3302 – 216 Mhz	16 MB	BCM3302	Linux OpenWRT	wireless utils	192.168.1.1

Tabla 4.2 Descripción de la Infraestructura Utilizada.

Como se aprecia en Figura 1, el servidor de VoIP se encuentra conectado por medio cableado al punto de acceso y no conectado inalámbricamente, pues es el caso típico de una implementación real, ya que es poco frecuente que dos estaciones asociadas al mismo punto de acceso requieran una comunicación de voz entre ellas. Las llamadas generalmente van hacia otra red, o a la misma red pero a un equipo asociado a otro punto de acceso.

En cuanto a los dispositivos inalámbricos, se contó con equipos 802.11b y 802.11g. Los dispositivos usados fueron tarjetas inalámbricas PCI, PCMCIA, miniPCI e incluso puntos de acceso configurados como *wireless client*, obteniendo un ambiente totalmente heterogéneo con diferentes marcas y tipos de equipos. En cuanto al sistema operativo, se

instaló GNU/Linux en todos los equipos, tanto cliente como servidor, e incluso en uno de los puntos de acceso, lo que dio una flexibilidad enorme en configuraciones, toma de datos y permitió el uso de una gran cantidad de herramientas como Iptraf para el diagnóstico de interfaces de red, ITG para la generación de Tráfico de Internet y Asterisk como Servidor de VoIP, los cuales se detallarán a continuación.

4.2.2 Métodos de Evaluación

Debido a que en la literatura encontrada, no existe un método para establecer capacidad de canales de voz en una red de conmutación de paquetes, se establece un procedimiento empírico descrito a continuación:

1. Se verifican condiciones iniciales adecuadas del sistema: tráfico actual, ancho de banda disponible, nivel de interferencias.
2. Se realiza la generación de llamadas una por una, vigilando constantemente los valores de parámetros como jitter, retardo y pérdidas, seguido de una evaluación subjetiva mediante la apreciación de la calidad de la voz en las nuevas llamadas generadas.
3. En el momento en que uno de los parámetros alcanza un valor mayor a los sugeridos en la sección 4.1, se determina que se ha llegado al tope de la capacidad.
4. Este procedimiento se repite para cada tipo de codificación, obteniendo la capacidad en diferentes condiciones de operación.

Para asegurar la validez de los resultados, el procedimiento se debe hacer de por lo menos dos formas distintas, buscando una validación recíproca de resultados.

Para la evaluación se usaron dos herramientas “open source” encargadas de generar las llamadas. Una fue el Generador de Tráfico de Internet – ITG [13], la otra fue el Servidor de VoIP Asterisk® [10]. Se usaron solo las codificaciones que estuvieran presentes tanto en ITG como en Asterisk, como se pueden apreciar en la Tabla 4.3 para ITG y en la Tabla 4.4 para Asterisk.

Códec	G.711	G.723.1	G.729
Tasa de Bits (Kbps)	84	16.6	28
Paquetes /segundo	50	26	50

Tabla 4.3 Atributos de los codecs de ITG

Códec	G.711	G.723.1	G.729
Tasa de Bits (Kbps)	84	18.2	28
Paquetes /segundo	50	26	50

Tabla 4.4 Atributos de los codecs de Asterisk

En el anexo A. Se puede encontrar en detalle como se realizó la implementación de las dos herramientas, incluyendo la configuración y scripts generadores de llamadas, configuración de servidores y clientes

4.2.3 Despliegue de resultados

4.2.3.1 Resultados obtenidos del trabajo con ITG

En primer lugar se tomó el dato de utilización de ancho de banda y cantidad de paquetes por segundo de cada códec por medio de IPtraf. Los resultados coincidieron exactamente con los que muestra la tabla II, como se puede verificar en las figuras de la 4.4 a la 4.6.

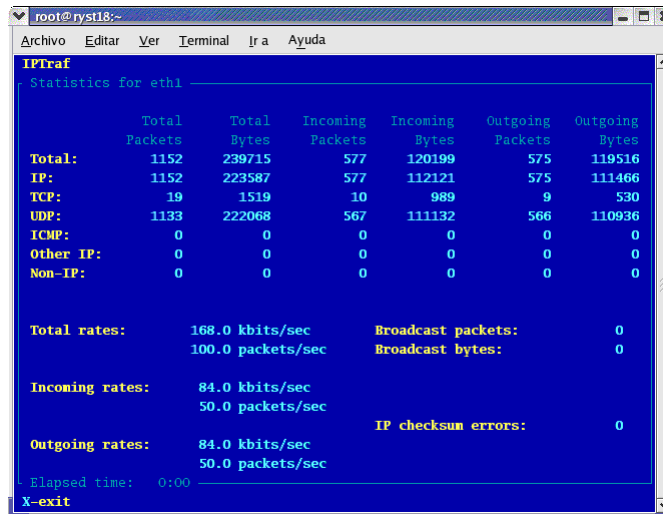


Figura 4.4 Datos de Iptraf de una llamada con G.711

```

root@ryst18:~# IPTraf
Statistics for eth1

Total      Total      Incoming   Incoming   Outgoing   Outgoing
Packets    Bytes      Packets    Bytes      Packets    Bytes
Total:     954        67219     478        33881     476        33338
IP:        954        53863     478        27189     476        26674
TCP:       23         1727      12         1093      11         634
UDP:       931        52136     466        26096     465        26040
ICMP:      0          0         0          0         0          0
Other IP:  0          0         0          0         0          0
Non-IP:    0          0         0          0         0          0

Total rates:      56.0 kbits/sec      Broadcast packets: 0
                  100.0 packets/sec      Broadcast bytes: 0

Incoming rates:   28.0 kbits/sec
                  50.0 packets/sec

Outgoing rates:   28.0 kbits/sec
                  50.0 packets/sec

IP checksum errors: 0

Elapsed time: 0:00
X-exit

```

Figura 4.5 Datos de Iptraf de una llamada con G.729

Luego de esto se inició la generación de llamadas en la forma explicada anteriormente; con G.711 se logró realizar hasta 11 llamadas, con G.729 hasta 14 y con G.723 hasta 28. Luego, se procedió a decodificar los archivos de registro (Logs) obtenidos con cada uno de los códecs: G.711, G.729 y G.723.

```

root@ryst18:~# IPTraf
Statistics for eth1

Total      Total      Incoming   Incoming   Outgoing   Outgoing
Packets    Bytes      Packets    Bytes      Packets    Bytes
Total:     602        48541     301        24513     301        24028
IP:        602        40113     301        20299     301        19814
TCP:       29         2295      14         1357      15         938
UDP:       573        37818     287        18942     286        18876
ICMP:      0          0         0          0         0          0
Other IP:  0          0         0          0         0          0
Non-IP:    0          0         0          0         0          0

Total rates:      33.3 kbits/sec      Broadcast packets: 0
                  52.0 packets/sec      Broadcast bytes: 0

Incoming rates:   16.6 kbits/sec
                  26.0 packets/sec

Outgoing rates:   16.6 kbits/sec
                  26.0 packets/sec

IP checksum errors: 0

Elapsed time: 0:00
X-exit

```

Figura 4.6 Datos de Iptraf de una llamada con G.723

Para esto se utiliza el Decodificador del ITG, la sintaxis es:

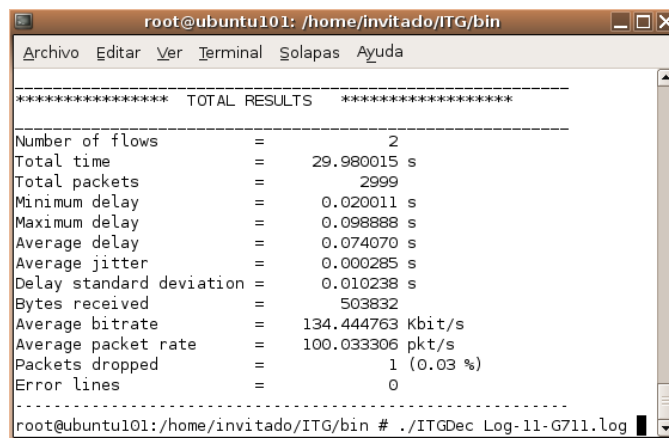
```
./ITGDec [log]
```

En este caso los logs siguieron la siguiente convención de nombrado: Log-[cantidad del llamadas]-[códec].log. Por ejemplo, para el caso de la decodificación de 11 llamadas

realizadas con G.711 este comando quedaría:

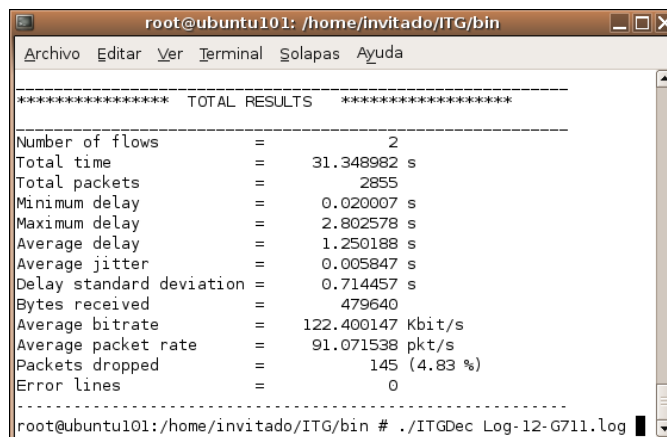
```
./ITGDec Log-11-G711.log
```

Los siguientes pares de figuras muestran los resultados obtenidos sobre la base de los cuales se estableció el límite de capacidad de acuerdo a cada códec. En la figura 4,7 que aparece primero las condiciones de retardo, *jitter* y pérdida de paquetes no superan los límites para una óptima comunicación de VoIP pero en la figura 4.8 (que es cuando se aumenta una llamada más) los supera considerablemente. En la parte inferior de cada una aparece el comando utilizado para la decodificación de los resultados.



```
root@ubuntu101: /home/invitado/ITG/bin
-----
***** TOTAL RESULTS *****
-----
Number of flows      =          2
Total time           =    29.980015 s
Total packets        =    2999
Minimum delay        =    0.020011 s
Maximum delay        =    0.098888 s
Average delay        =    0.074070 s
Average jitter       =    0.000285 s
Delay standard deviation = 0.010238 s
Bytes received       =    503832
Average bitrate      =   134.444763 Kbit/s
Average packet rate  =   100.033306 pkt/s
Packets dropped      =          1 (0.03 %)
Error lines          =          0
-----
root@ubuntu101:/home/invitado/ITG/bin # ./ITGDec Log-11-G711.log
```

Figura 4.7 Decodificación de 11 llamadas con G.711



```
root@ubuntu101: /home/invitado/ITG/bin
-----
***** TOTAL RESULTS *****
-----
Number of flows      =          2
Total time           =   31.348982 s
Total packets        =    2855
Minimum delay        =    0.020007 s
Maximum delay        =    2.802578 s
Average delay        =    1.250188 s
Average jitter       =    0.005847 s
Delay standard deviation = 0.714457 s
Bytes received       =    479640
Average bitrate      =   122.400147 Kbit/s
Average packet rate  =    91.071538 pkt/s
Packets dropped      =     145 (4.83 %)
Error lines          =          0
-----
root@ubuntu101:/home/invitado/ITG/bin # ./ITGDec Log-12-G711.log
```

Figura 4.8. Decodificación de 12 llamadas con G.711

```

root@ubuntu101: /home/invitado/ITG/bin
-----
***** TOTAL RESULTS *****
-----
Number of flows      =          2
Total time           =    30.019855 s
Total packets        =         2999
Minimum delay        =     0.020010 s
Maximum delay        =     0.322185 s
Average delay        =     0.068664 s
Average jitter       =     0.000695 s
Delay standard deviation = 0.016357 s
Bytes received       =         83972
Average bitrate      =    22.377723 Kbit/s
Average packet rate  =    99.900549 pkt/s
Packets dropped      =          1 (0.03 %)
Error lines          =          0
-----
root@ubuntu101:/home/invitado/ITG/bin # ./ITGDec Log-14-G729.log

```

Figura 4.9 Decodificación de 14 llamadas con G.729

```

root@ubuntu101: /home/invitado/ITG/bin
-----
***** TOTAL RESULTS *****
-----
Number of flows      =          3
Total time           =    30.426446 s
Total packets        =         4271
Minimum delay        =     0.020006 s
Maximum delay        =     2.249148 s
Average delay        =     1.014241 s
Average jitter       =     0.004659 s
Delay standard deviation = 0.551907 s
Bytes received       =        119588
Average bitrate      =    31.443173 Kbit/s
Average packet rate  =   140.371307 pkt/s
Packets dropped      =         229 (5.09 %)
Error lines          =          0
-----
root@ubuntu101:/home/invitado/ITG/bin # ./ITGDec Log-15-G729.log

```

Figura 4.10 Decodificación de 15 llamadas con G.729

```

root@ubuntu101: /home/invitado/ITG/bin
-----
***** TOTAL RESULTS *****
-----
Number of flows      =          5
Total time           =    30.062761 s
Total packets        =         3879
Minimum delay        =     0.076936 s
Maximum delay        =     2.426430 s
Average delay        =     0.904287 s
Average jitter       =     0.007613 s
Delay standard deviation = 0.482350 s
Bytes received       =        147402
Average bitrate      =    39.225140 Kbit/s
Average packet rate  =   129.030065 pkt/s
Packets dropped      =          21 (0.54 %)
Error lines          =          0
-----
root@ubuntu101:/home/invitado/ITG/bin # ./ITGDec Log-28-G723.log

```

Figura 4.11 Decodificación de 28 llamadas con G.723

```
root@ubuntu101: /home/invitado/ITG/bin
-----
***** TOTAL RESULTS *****
-----
Number of flows      =          5
Total time           =    30.107248 s
Total packets        =         3720
Minimum delay        =     0.038467 s
Maximum delay        =     1.852169 s
Average delay        =     1.210702 s
Average jitter       =     0.007074 s
Delay standard deviation =  0.498373 s
Bytes received       =        141360
Average bitrate      =    37.561719 Kbit/s
Average packet rate  =    123.558287 pkt/s
Packets dropped       =          180 (4.62 %)
Error lines          =              0
-----
root@ubuntu101: /home/invitado/ITG/bin #
```

Figura 4.12 Decodificación de 29 llamadas con G.723

Estos resultados son los que dan soporte a las capacidades encontradas, esto es: para G.711 hasta 11 llamadas, para G.729 hasta 14 llamadas y para G.723 hasta 28 llamadas, lo anterior a razón de que siempre que se aumentaba *una sola* llamada, los límites de retardo (delay) y pérdida de paquetes (packets dropped) eran sobrepasados “notoriamente”, lo cual se aprecia en figuras 4.8, 4.10 y 4.12 para G.711, G.729 y G.723 respectivamente, especialmente para la pérdida de paquetes (*Packets dropped*).

4.2.3.2 Resultados obtenidos del trabajo con Asterisk

Inicialmente se tomaron los datos de consumo de ancho de banda para una sola llamada con cada uno de los 3 códecs, obteniendo resultados similares a los obtenidos con ITG:

G.711	82.4 Kbits/sg
G.729	27.8Kbits/sg
G.723	18.3Kbits/sg

Es importante anotar también que se tomaron medidas para varias llamadas y el consumo de ancho de banda fue exactamente el consumo de una sola llamada multiplicado por el número de llamadas, tal como se observa en las figuras 4.13 y 4.14, además, se consumió el mismo ancho de banda cuando se hicieron llamadas desde diferentes clientes que cuando se hicieron llamadas desde un solo cliente.

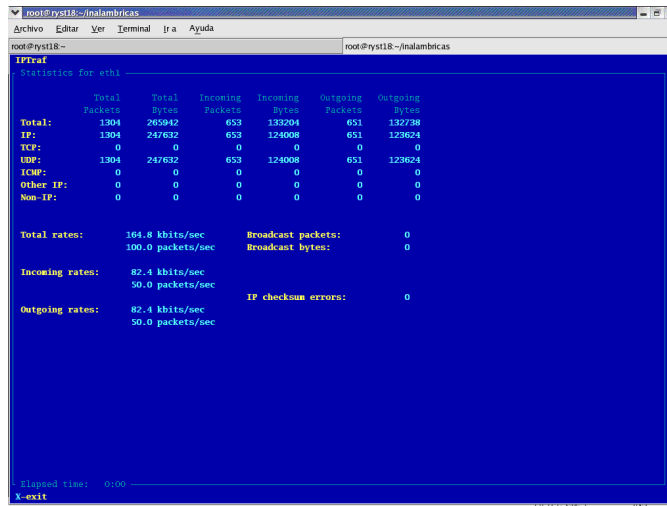


Figura 4.13 Datos de Iptraf con una llamada G.711

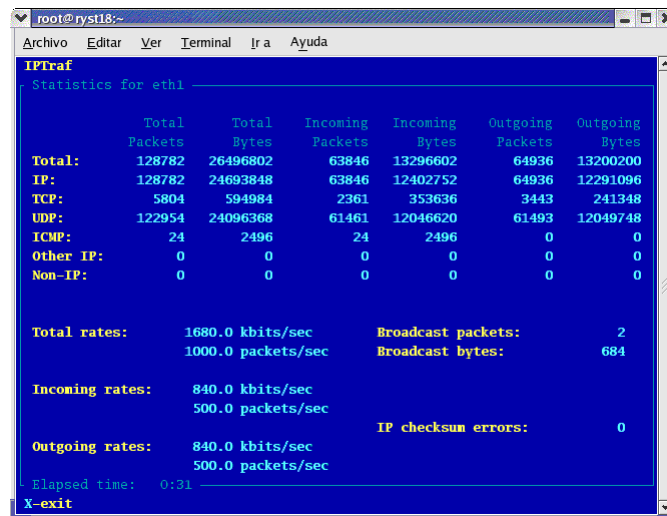


Figura 4.14 Datos de Iptraf con diez llamadas G.711

Para hacer una evaluación de los resultados obtenidos, se tomaron datos empezando por una llamada y terminando en el momento en que la cantidad de llamadas no permitieran una comunicación fluida.

4.2.3.2.1 Resultados con G.711

El tipo de medidas que se tomaron se pueden observar en la figura 4.15 que muestra los

resultados tabulados a medida que se iban generando nuevas llamadas.

```

ryst18*CLI> iax2 show channels
Channel Peer Username ID (Lo/Rem) Seq (Tx/Rx) Lag Jitter JitBuf Format
IAX2/cliente@client 192.168.1.104 cliente6 00001/00003 00137/00136 00004ms 0000ms 0000ms ulaw
IAX2/cliente@client 192.168.1.100 cliente6 00003/00004 00053/00052 00006ms 0003ms 0000ms ulaw
IAX2/cliente@client 192.168.1.100 cliente6 00007/00005 00053/00052 00006ms 0003ms 0000ms ulaw
IAX2/cliente@client 192.168.1.101 cliente6 00010/00002 00014/00013 00004ms 0003ms 0000ms ulaw
IAX2/cliente@client 192.168.1.103 cliente6 00013/00004 00078/00077 00006ms 0001ms 0000ms ulaw
IAX2/cliente@client 192.168.1.102 cliente6 00016/00004 00036/00035 00003ms 0003ms 0000ms ulaw
IAX2/cliente@client 192.168.1.102 cliente6 00018/00005 00034/00033 00006ms 0008ms 0000ms ulaw
IAX2/cliente@client 192.168.1.105 cliente6 00020/00003 00005/00004 00004ms 0002ms 0000ms ulaw
IAX2/cliente@client 192.168.1.103 cliente6 00021/00005 00078/00077 00006ms 0002ms 0000ms ulaw
IAX2/cliente@client 192.168.1.105 cliente6 00025/00007 00005/00004 00008ms 0003ms 0000ms ulaw
10 active IAX channel(s)

```

Figura 4.15 Conexiones, Retardo, Jitter y Códec de 10 llamadas en curso sobre el servidor de VoIP

Como solo se contó con 6 clientes, a partir de la 7 llamada fue necesario generar más llamadas por cada cliente.

Para el códec G711 se tuvo un comportamiento adecuado y se obtuvieron buenos resultados hasta la llamada número 12. Con la llamada número 13, el retardo y *jitter* aumentó drásticamente y se pudo percibir una degradación la calidad.

En la figura 4.16 se observa la degradación del *jitter* a medida que se aumentan las llamadas

En la figura 4.17 se observa la degradación del retardo a medida que se aumentan las llamadas

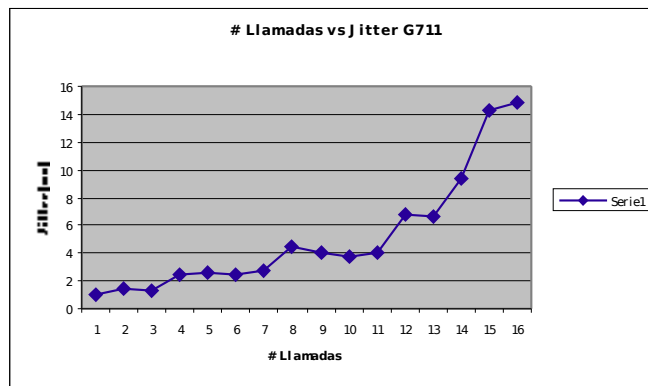


Figura 4.16 Observación del *Jitter* a medida que las llamadas aumentan.

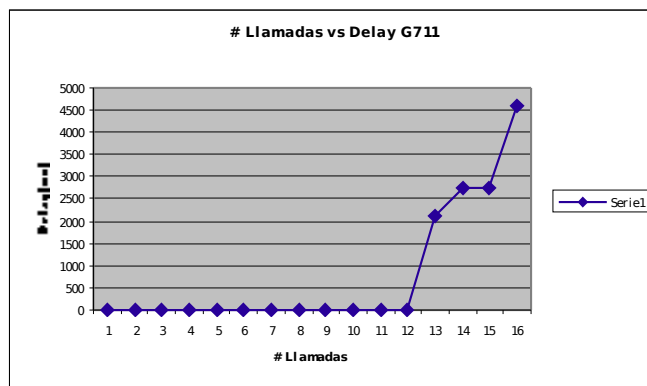


Figura 4.17 Observación del Retardo a medida que las llamadas aumentan.

No Llamadas	1	2	3	4	5	6	7	8
Jitter	1	1,5	1,33	2,5	2,6	2,5	2,71	4,5
Delay	3	2	2,67	3	3,8	4,67	4,57	5,13
No Llamadas	9	10	11	12	13	14	15	16
Jitter	4	3,7	4	6,75	6,62	9,36	14,26	14,87
Delay	4,5	4,5	5,64	9,17	2130	2751	2734	4588

Tabla 4.5 Resumen de Resultados G711

Como se ve en las figuras 19 y 20 y en la tabla IV, el *jitter* y retardo son muy buenos (están dentro de los límites) hasta la llamada 12, pero en el momento en que se realiza la llamada siguiente el retardo aumenta dramáticamente a pesar de que el *jitter* se mantiene aun apto para una comunicación adecuada de voz. Para verificar lo anterior, se puso un *softphone* con parlantes para apreciar la calidad de la comunicación obteniendo las siguientes anotaciones:

Llamada 11: El sonido es de excelente calidad, no se aprecian entrecortes o chasquidos.

Llamada 12: El sonido continúa siendo excelente, aunque se alcanzan a apreciar algunos pequeños chasquidos muy esporádicos, pero nada que degrade la comunicación.

Llamada 13: La calidad se degrada apreciablemente, se aprecian fácilmente chasquidos y algunos entrecortes, toma mucho tiempo el establecimiento de una nueva conversación.

Llamada 14: Se hace más evidente la pérdida de calidad, se aprecian muchos entrecortes y se dificulta el inicio de nuevas sesiones.

Llamada 15: No se entienden muchas partes de la conversación, se entrecorta

constantemente y por largos periodos de tiempos.

Llamada 16: Después de muchos intentos, se logra establecer la comunicación, pero no es entendible.

Llamada 17: Cada vez que se intentó iniciar la llamada 17, se cayó la llamada, y “tumbó” más de la mitad de las que se estaban cursando.

Es importante anotar, que se repitió el experimento para verificar que la información obtenida es correcta, y que se obtuvieron los mismos datos en las dos ocasiones.

El proceso para obtener los resultados con los códecs G.729 y G.723, fue el mismo que se siguió para G.711. Los resultados obtenidos se muestran en las gráficas de las figuras 4.18 y 4.19:

4.2.3.2.2 Resultados con G.729

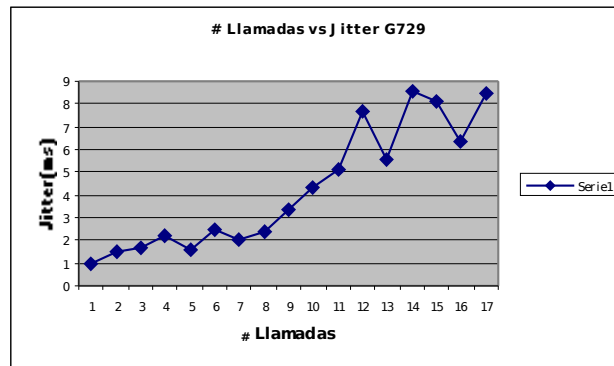


Figura 4.18 Observación del Jitter a medida que las llamadas aumentan con G729.

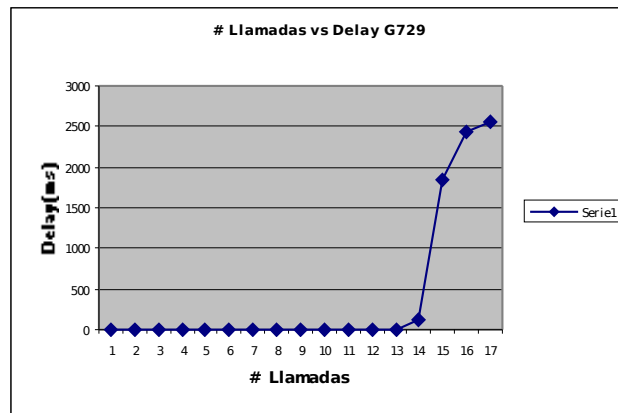


Figura 4.19 Observación del Retardo a medida que las llamadas aumentan con G729

Como se observa en las gráficas, el comportamiento es similar al que tienen las llamadas con G711, pero con una capacidad mayor de llamadas. Como se observa, hasta la llamada 14 el funcionamiento es estable y con buena calidad, pero en la llamada 15, el retardo aumenta a tal punto, que se afecta la calidad de todas las llamadas y se empiezan a sentir entrecortes en la comunicación. A partir de este punto y hasta la llamada 17 se degrada más y más la calidad de la llamada hasta no ser entendible y/o estable. Cuando se intenta generar la llamada 18, no se puede establecer, e incluso tumba la mayoría de las llamadas que se están cursando. Este resultado muestra que a pesar de que el ancho de banda que consume G729 es la tercera parte de G711, la capacidad de llamadas solo aumento en un 14%, lo que permite ver que la mayor limitante en las comunicaciones es la contienda por el medio antes que el consumo de ancho de banda

4.2.3.2.3 Resultados con G.723

Como se observa en las figuras 4.20 y 4.21, con el códec G.723, se tiene un comportamiento similar a los dos anteriores casos, pero con una capacidad de llamadas bastante superior.

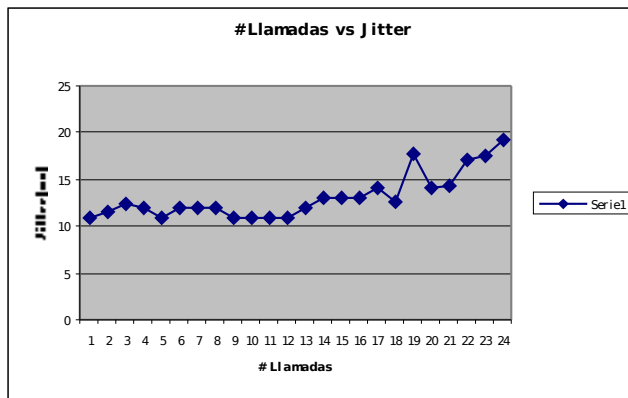


Figura 4.20 Observación del Jitter a medida que las llamadas aumentan con G723

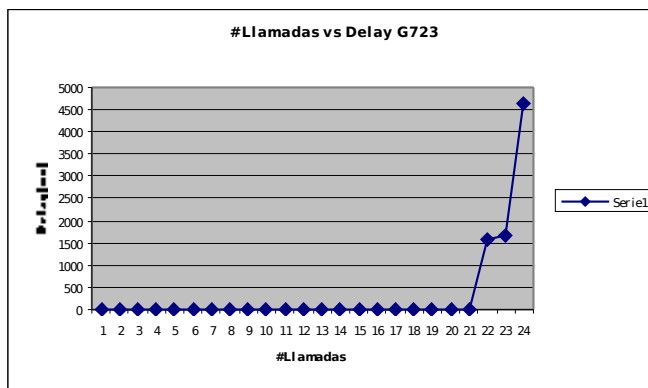


Figura 4.21 Observación del Retardo a medida que las llamadas aumentan con G723

Con el códec G723, se logran cursar 21 llamadas antes que empiecen a presentarse pérdidas considerables en la calidad. A partir de la llamada 22, se empieza a degradar la calidad y en llamada 24 la comunicación es bastante difícil, al tratar de establecer la llamada 25, se desconectan muchas de las llamadas que se estuvieran cursando.

Se deben observar dos comportamientos recurrentes en las pruebas hechas con los 3 códecs:

1. El aumento de *jitter* se hace de un modo casi lineal, tiene una pendiente pequeña y no sobrepasa los límites para una buena conversación de voz, el retardo es muy pequeño y cercano a cero hasta cierto punto, pero después de éste tiene un cambio abrupto que pasa de alrededor de 5 ms a varios segundos.

2. Cuando se trató de generar nuevas llamadas en el momento que se contaba con un retardo excesivo, se ocasionó la pérdida de conexión.

4.2.3.3 Análisis de resultados

El desarrollo de este estudio de capacidad por métodos experimentales arroja una serie de aspectos importantes para el trabajo con VoIP en redes inalámbricas que se anotan a continuación:

Las comunicaciones de VoIP en la redes inalámbricas tienen límites abruptos, esto quiere decir que luego que se supera el límite de capacidad, una comunicación siguiente queda sin ninguna probabilidad de ser soportada a razón del retardo y pérdida de paquetes tan significativo que se presenta. Por este motivo es necesario establecer nuevos criterios de diseño con VoIP pues los tradicionalmente utilizados para Ethernet cableada, donde la capacidad es directamente proporcional a la velocidad o ancho de banda disponible, no son aplicables.

Los parámetros que más incidencia tienen en el límite del número de llamadas concurrentes en un sistema de VoWLAN son el retardo y la pérdida de paquetes que tiene un cambio abrupto cuando se supera una determinada cantidad de llamadas. En cambio el *jitter* presenta un comportamiento lineal, con pequeña pendiente, que se mantiene fácilmente dentro de los límites permitidos para una buena conversación.

Es muy importante en un sistema de VoWLAN, poner un límite en el número de llamadas que puedan establecerse por cada punto de acceso, pues si se supera este límite, se ocasionará no solo una pérdida de calidad, sino la desconexión de las llamadas que se estén cursando a través del punto de acceso.

Debe tenerse en cuenta también el consumo de máquina que puede tenerse con la codificación de voz al usar G.729 o G.723, ya que en dispositivos móviles con poca

capacidad de procesamiento puede no ser posible usar un códec diferente a G.711.

El deterioro en la calidad cuando el retardo aumenta abruptamente es fácilmente comprobable. Cada vez que se superaba el umbral de llamadas encontrado para cada códec, se escuchaba una notoria disminución de la calidad en la comunicación a través de los parlantes conectados al *softphone*, en donde se escuchaban entrecortes y múltiples ruidos.

Mientras en las redes cableadas el tipo de códec que se utiliza determina casi linealmente la capacidad (cantidad de llamadas) que puede tener el sistema, en las redes inalámbricas aunque un códec con menor consumo de ancho de banda permite un mayor número de llamadas, este efecto no es tan significativo como se notó en los experimentos donde se comprobó que lo más crítico venía a ser la contienda por el medio.

Los resultados obtenidos de llevar a cabo las pruebas con ITG y Asterisk sirven para un proceso de validación implícita del laboratorio realizado. La única diferencia se obtuvo con el códec G.723 lo cual fue consecuencia directa de tener características deferentes para las dos aplicaciones. Pero fue muy interesante ver que tanto para G.711 y G.729 los resultados fueron los mismos.

Con las implementaciones futuras de los sistemas de calidad de servicio, es probable que al ser la voz el servicio privilegiado, lo que se verá afectado será el desempeño en cuanto a la transferencia de datos tipo *best effort*, sobre todo si los fabricantes hacen como con 802.11 donde solamente implementaron el esquema DCF, y no se tenía un elemento central que controlara el acceso.

Este trabajo responde muchas de las preguntas planteadas al inicio del mismo, pero plantea algunas nuevas que serán respondidas con los siguientes trabajos mostrados en el capítulo: la congestión en este tipo de redes se ve que es proporcional al número de paquetes por segundo más que por el ancho de banda que consumen los paquetes (entre otras los paquetes de voz son pequeños pero requieren ser despachados

inmediatamente). Esto indica que sería interesante estudiar el impacto de diferentes tamaños de paquetes. Por ejemplo, colocar dos tramas en un paquete y por tanto bajar la tasa de paquetes a la mitad (reduciendo por tanto el *overhead* del paquete también a la mitad) lo que sin duda reducirá la congestión y se podrían esperar mejor desempeño de la red. Sin embargo, el inconveniente es que el retardo por procesamiento aumentaría y la sensibilidad a la pérdida de paquetes también, luego vendrían algunas preguntas inmediatas que se buscará responder en la siguiente sección ¿La disminución de paquetes perdidos debido al mejor desempeño de la red compensará el aumento en la sensibilidad a los paquetes que se puedan perder? ¿Qué es más conveniente optimizar en la práctica?

4.3 LINEAMIENTOS PARA LA SELECCIÓN DE PARÁMETROS PARA COMUNICACIONES DE VOIP SOBRE 802.11

El modo tradicional de realizar las comunicaciones de voz en el mundo y que se basaba en la conmutación de circuitos, se ha venido reemplazando gradualmente desde hace varios años a conmutación de paquetes y principalmente a la transmisión de la voz a través del protocolo de internet IP, debido a factores económicos y tecnológicos atractivos para las empresas de telecomunicaciones, entre ellos: un rápido despliegue sobre redes ip ya existentes, la posibilidad de ofrecer servicios de voz y datos sobre el mismo medio y la facilidad para la ampliación de la cobertura. Aparentemente, el uso de la tecnología de VoIP es independiente de las características del medio por el que se transmite y de los protocolos de nivel 2 que se usan para la transmisión de los datos por lo que la voz podría ser indistintamente transmitida en modo de datos a través de enlaces de cobre, fibra, inalámbricos, etc y usando Ethernet, HDLC, WiFi, ATM, entre otros. En la práctica, se observa que existen diferentes fenómenos que afectan notablemente la calidad del servicio y que bajo ciertas condiciones, como: enlaces con pérdidas de paquetes, anchos de banda pequeños o altas latencias, se hace inviable el establecimiento de una

comunicación dentro de un mínimo de calidad. En la presente sección se muestran las diferentes alternativas, técnicas y parámetros a tener en cuenta para lograr establecer una comunicación con una adecuada percepción del usuario a pesar de que la misma no presente las mejores condiciones. Se explican los diferentes tipos de codificaciones, la incidencia que tiene el tipo de muestreo y el tamaño de la carga útil en cada paquete IP sobre la calidad de la comunicación, el efecto de la pérdida de paquetes, el papel del tamaño de la cabecera en el aumento del ancho de banda requerido y en general de los principales parámetros que pueden incidir en la calidad de una llamada usando VoIP. Al finalizar se hace un compendio de cómo establecer la comunicación de VoIP, optimizando los parámetros mencionados, realizando una adecuada escogencia de la codificación y muestreo, haciendo uso de algoritmos para el relleno en la pérdida de paquetes, compresión de cabeceras, etc, logrando mejorar notablemente la calidad de la misma, aun en presencia de los fenómenos nocivos anteriormente comentados. Por último, se realiza la comprobación en laboratorio, cuyos resultados validan el trabajo realizado.

4.3.1 Selección de los mejores parámetros en redes no idóneas para transmisión de VoIP

No existe una combinación de parámetros que sea la idónea para todos los tipos de redes. A continuación se mostrarán escenarios que ilustran diferentes enlaces en condiciones poco óptimas y como pueden optimizarse los parámetros para que pueda establecerse una comunicación de VoIP de manera adecuada.

4.3.1.1 Enlaces de Ancho de Banda limitado

Este es uno de los casos más frecuentes. Accesos a sitios remotos, de difícil acceso o con redes antiguas usualmente disponen de anchos de banda limitados, o en los que un acceso a internet de mayor velocidad puede ser muy costoso. En esta categoría también se incluyen los casos en que se dispone de un enlace de datos de buena velocidad, pero se requiere realizar varias llamadas concurrentes sobre el mismo enlace.

Tres aspectos son claves para lograr el ahorro del ancho de banda: la selección del

códec, el tamaño de la carga útil y la compresión de las cabeceras.

En cuanto al ancho de banda, se recomienda usar un códec de alta compresión como el g.723.1, anotando que este códec realiza muestras de 30ms por lo que el retardo puede ser un poco mayor que con otros.

Respecto al tamaño de la carga útil se recomienda usar solo una muestra por paquete, pues, la muestra del códec g.723.1 ya es lo suficientemente grande y usar dos o más muestras por paquete generaría un retardo excesivo afectando notablemente la percepción del servicio.

La mayor reducción de ancho de banda se logra al realizar compresión en las cabeceras. Donde sea posible, se recomienda usar alguno de los algoritmos mencionados anteriormente.

Adicionalmente, si el ancho de banda disponible es excesivamente pequeño, se puede usar algoritmos VAD para disminuir aun más el ancho de banda por canal.

En la tabla 4.6, se muestra el ancho de banda por canal de voz usando diferentes códec, y el efecto de añadir compresión a las cabeceras usando cRTP

Codec	BW (Kbps)	BW usando cRTP (Kbps)
G.711 (64 Kbps)	82.8 Kbps	67.6 Kbps
G.729 (8 Kbps)	26.8 Kbps	11.6 Kbps
G.723.1 (6.3 Kbps)	18.9 Kbps	8.8 Kbps
G.723.1 (5.3 Kbps)	17.9 Kbps	7.7 Kbps
G.726 (32 Kbps)	50.8 Kbps	35.6 Kbps
G.726 (24 Kbps)	42.8 Kbps	27.6 Kbps
G.728 (16 Kbps)	28.5 Kbps	18.4 Kbps

Tabla 4.6 Ancho de Banda por codec con y sin compresión de cabeceras

Se puede observar que es posible realizar una comunicación de VoIP usando los parámetros adecuados aún en una conexión de 9600bps disponibles en lugares remotos, por antiguos radioenlaces.

4.3.1.2 Enlaces con Altos Retardos

En algunos escenarios, el ancho de banda disponible no supone ser un problema, pero si lo es el alto retardo extremo a extremo presente en enlaces satelitales o enlaces con muchos saltos. En estos casos, los parámetros más importantes son la selección del códec y el tamaño de muestras a incluir en la carga útil, pues se debe buscar como objetivo, añadir el mínimo retardo posible a la comunicación.

Para estos casos se recomienda usar un códec con el menor retardo posible como g.711, g.726 o g.728 y adicionalmente enviar solo una muestra como carga útil. Debido a que este tipo de enlaces presentan un retardo apreciable, se recomienda la implementación de un método de cancelación de eco y un *jitter buffer* de tamaño medio que alcance a compensar las variaciones entre llegadas de los paquetes y que por otro lado no añada mucho retardo al ya existente.

En la tabla 4.7 se aprecia el retardo inducido por diferentes códecs.

Codec	Retardo Cod/Dec	Retardo Look-Ahead	Retardo Total
G.711	0.125ms	0ms	0.25ms
G.729	10ms	5ms	25ms
G.723.1	30ms	7.5ms	67.5ms
G.726	0.125ms	0ms	0.25ms
G.728	0.625ms	0ms	1.25ms

Tabla 4.7. Retardos en cada códec, para una sola muestra por paquete

4.3.1.3 Enlaces con Pérdidas de paquetes constantes

En algunos tipos de enlaces, como inalámbricos, pares de cobre en condiciones no óptimas, o en redes saturadas, es común la pérdida esporádica de algunos paquetes. El tráfico de VoIP se transporta generalmente con RTP/UDP, esto lo hace sensible a la pérdida de paquetes en el camino, los cuales dependiendo del tipo de codificación y tamaño de carga útil, pueden ser perceptibles y afectar la calidad de la comunicación. Para este tipo de enlaces, se recomienda el uso de codificaciones con muestras pequeñas y de solo una muestra por paquete. Adicionalmente se debe usar un algoritmo para el relleno o compensación de los paquetes perdidos PLC. Debe tenerse en cuenta que las técnicas PLC son altamente efectivas solo con pérdidas ocasionales y de no más

de 30ms en total. En casos en que las pérdidas son más frecuentes y/o más continuas, la técnica aun ayuda a mejorar la percepción, pero no garantiza inteligibilidad de la comunicación.

Algunas técnicas que pueden ser usadas son:

1. Inserción de Ceros
2. Repetición de la última muestra recibida
3. Predicción de la muestra perdida

4.3.2 Comprobación y validación en laboratorio

Para realizar el montaje de laboratorio, se recurrió a varios elementos para poder realizar las comunicaciones de VoIP y recrear correctamente los ambientes propuestos. Se realizó un montaje de 1 Servidor de VoIP y 6 clientes conectados via Fast Ethernet y via inalambrica 802.11g, se uso el simulador de tráfico D-ITG para la generación de tráfico de VoIP usando compresión de cabeceras cRTP, software como TC para la generación de pérdida de paquetes y limitación del ancho de banda, IPTRAF para realizar las mediciones de tráfico y Asterisk como servidor de VoIP. En cuanto a codecs, se trabajo con los codecs ITU disponibles para Asterisk g.711, g.729 y g.723. Debido a que estos codecs no tienen implementado PLC, se uso SPEEX para realizar la evaluación de percepción ante una comunicación con pérdidas con/sin PLC.

Para la evaluación audible, se uso el Softphone X-lite y un adaptador telefónico ATA Sipura SPA-2000.

En el laboratorio se recrearon 3 escenarios en los que se puede observar claramente como el uso de codecs adecuados, compresión de cabeceras, cancelación de eco, algoritmo PLC, entre otros, permiten una comunicación adecuada a pesar de las bajas prestaciones de los enlaces observados en cada escenario.

4.3.2.1 Escenario 1: Canal de 28.800 bps, pérdidas del 1%, retardo de 100ms (enlace conmutado V.34)

Simulación de condiciones del canal usando TC:

```
tc qdisc add dev eth0 root netem delay 100ms
```

```
tc qdisc add dev eth0 root netem loss 1%
```

```
tc class add dev eth0 parent 1: classid 1:1 htb rate 28kbps ceil 28kbps
```

Debido a que el retardo es manejable se puede usar un códec de retardo promedio, sin embargo es necesaria la escogencia de un códec que consuma poco ancho de banda. observando las condiciones del escenario, se escoge g.729. Como se observa en la Figura 4.22, el tráfico es inferior al limite puesto con tc (27,8 kbps) y se puede realizar la comunicación de voz con una buena percepción, totalmente inteligible y un retardo manejable.

	Total Packets	Total Bytes	Incoming Packets	Incoming Bytes	Outgoing Packets	Outgoing Bytes
Total:	10832	751155	5415	376467	5417	374688
IP:	10832	598993	5418	300297	5420	299006
TCP:	52	6799	23	3397	29	3402
UDP:	10742	572880	5357	277596	5391	295604
ICMP:	38	19304	38	19304	0	0
Other IP:	0	0	0	0	0	0
Non-IP:	0	0	0	0	0	0
Total rates:		55.5 kbits/sec 102.8 packets/sec		Broadcast packets:	0	Broadcast bytes: 0
Incoming rates:		27.8 kbits/sec 51.4 packets/sec				
Outgoing rates:		27.7 kbits/sec 51.4 packets/sec		IP checksum errors:	0	

Figura 4.22 Consumo BW G.729

Para realizar la prueba usando cRTP, se uso D-ITG, pues Asterisk no tiene implementada esta característica. Es importante mencionar que cRTP es solo implementable cuando en todos los puntos del trayecto de la comunicación se dispone de equipos que cumplan con el protocolo, por lo cual en varios ambientes no es viable su implementación. Usando cRTP se disminuyó el consumo de BW a menos del 50%, como se puede observar en la figura 4.23 que muestra el reporte generado por D-ITG usando cRTP.

```

-----
***** TOTAL RESULTS *****
-----
Number of flows      =          2
Total time           =    30.019855 s
Total packets       =         2999
Minimum delay       =    0.120010 s
Maximum delay       =    0.322185 s
Average delay       =    0.168664 s
Average jitter      =    0.000695 s
Delay standard deviation =    0.016357 s
Bytes received      =         83972
Average bitrate     =    11.377723 Kbit/s
Average packet rate =    99.900549 pkt/s
Packets dropped     =          1 (0.03 %)
Error lines         =          0
-----

```

Figura 4.23 Reporte D-ITG usando cRTP

4.3.2.2 Escenario 2: Canal de 128.000 bps, pérdidas del 1%, retardo de 300ms (enlace de datos con múltiples saltos)

Simulación de condiciones del canal usando TC:

```

tc qdisc add dev eth0 root netem delay 300ms
tc qdisc add dev eth0 root netem loss 1%
tc class add dev eth0 parent 1: classid 1:1 htb rate 128kbps ceil 128kbps

```

En este caso el enlace es lo suficientemente grande para soportar un códec de baja compresión, sin embargo requiere que el retardo adicional sea mínimo pues el retardo ya existente es bastante grande. Adicionalmente requiere de cancelación de eco. En la figura 4.24 se observa el consumo de BW de la comunicación. En la figura 4.25, se observa el retardo adicional generado por la codificación g.711 que es cercana a 0.

	Total Packets	Total Bytes	Incoming Packets	Incoming Bytes	Outgoing Packets	Outgoing Bytes
Total:	1304	265942	653	133204	651	132738
IP:	1304	247632	653	124008	651	123624
TCP:	0	0	0	0	0	0
UDP:	1304	247632	653	124008	651	123624
ICMP:	0	0	0	0	0	0
Other IP:	0	0	0	0	0	0
Non-IP:	0	0	0	0	0	0
Total rates:	164.8 kbits/sec 100.0 packets/sec		Broadcast packets:		0	
			Broadcast bytes:		0	
Incoming rates:	82.4 kbits/sec 50.0 packets/sec					
Outgoing rates:	82.4 kbits/sec 50.0 packets/sec		IP checksum errors:		0	

Figura 4.24 Consumo BW G.711

Bajo las condiciones planteadas, se percibió una comunicación de muy buena calidad, con un retardo perceptible pero manejable, la activación del cancelador de eco fue necesaria y se notó una clara mejoría al realizarlo.

ID (Lo/Rem)	Seq (Tx/Rx)	Lag	Jitter	JitBuf	Format
00007/00002	00009/00008	00006ms	0000ms	0000ms	ulaw
00007/00002	00011/00010	00003ms	0000ms	0000ms	ulaw
00007/00002	00011/00010	00003ms	0001ms	0000ms	ulaw
00007/00002	00011/00010	00003ms	0001ms	0000ms	ulaw

Figura 4.25 Retardo G.711

4.3.2.3 Escenario 3: Canal de 64.000 bps, pérdidas del 5%, retardo de 100ms (enlace de datos con múltiples saltos)

Simulación de condiciones del canal usando TC, parámetros:

```
tc qdisc add dev eth0 root netem delay 100ms
tc qdisc add dev eth0 root netem loss 5%
tc class add dev eth0 parent 1: classid 1:1 htb rate 64kbps ceil 64kbps
```

En este caso las pruebas fueron totalmente subjetivas pues se evaluó la percepción del usuario frente a las pérdidas de paquetes. Se usó codificación con SPEEX, pues con este códec se tiene la posibilidad de activar PLC. Con un 5% de paquetes perdidos, se percibieron claramente saltos y ruidos molestos en la comunicación, sin embargo al activar el relleno con PLC usando la última muestra recibida para rellenar los huecos se mitigó el efecto y se observaron mejoras en la percepción del usuario, demostrando la efectividad del uso de este tipo de técnicas.

4.3.3 Análisis de Resultados

Las tendencias muestran que en algún momento todo el tráfico de voz en el mundo se

hará por redes de conmutación de paquetes, por lo cual es de gran importancia el establecimiento de lineamientos para garantizar niveles adecuados de calidad de servicio sobre los diferentes medios y tipos de red.

Redes de datos que parecen ser no aptas para la transmisión de VoIP, como de baja velocidad, alta latencia o con pérdidas de paquetes, lo pueden ser, siempre y cuando se seleccionen los parámetros y técnicas adecuadas. Redes antiguas o de bajas prestaciones también pueden llegar a ser usadas.

Es muy importante conocer las características de la red de datos que se dispone para la transmisión de VoIP, no solo basta con disponer de grandes anchos de bandas como se suele suponer. En ocasiones este aspecto puede ser el menos importante. Aspectos como el retardo y el porcentaje de paquetes perdidos suelen tomar una mayor relevancia

4.4 CONSTRUCCIÓN Y DESARROLLO DE UN SISTEMA DE RECONOCIMIENTO DE VOZ EN ESPAÑOL

Como siguiente paso, se procedió al desarrollo y montaje de un sistema de reconocimiento de voz para el idioma español. Debido a la complejidad del mismo, se decidió por la construcción de un reconocedor de palabras aisladas pero se incluyó el entrenamiento necesario para hacerlo independiente del locutor. Aunque existen diferentes reconocedores comerciales o incluso de código abierto, se decidió por construir un reconocedor propio y así adaptarlo fácilmente al entorno en el que va a ser operado, que en este caso es un sistema de VoWLAN. A continuación, se describen el proceso mediante el cual se construyó el reconocedor.

4.4.1 Requerimientos

Como requerimientos para el reconocedor se tuvo:

- Implementar una interfaz oral diseñada para operar un sistema de telefonía IP carente de teclado.

- Desarrollar el reconocedor para su operación con palabras aisladas e independencia del locutor
- El reconocedor debe reconocer las frases que se muestran a continuación, para que pueda interactuar fácilmente con el sistema de comunicaciones de voz:
 - Consultar 'Nombre'
 - Llamar 'Nombre'
 - Marcar 'Número'
 - Correo
- El reconocedor deberá ser entrenado hasta tener una tasa de aciertos mínima del 90%
- Deberá reconocer efectivamente los dígitos del cero al nueve y nombres y apellidos personales comunes.

4.4.2 Implementación

En el capítulo 3 se muestran las diferentes alternativas para la implementación de un sistema de reconocimiento de voz, a partir del cual se seleccionaron los modelos ocultos de markov HMM, para tal efecto se uso el HTK (HMM Tool Kit). HTK consiste de un conjunto de herramientas, para trabajar con Modelos Ocultos de Markov, en especial, modelos relacionados con el reconocimiento automático del habla, por lo cual además provee herramientas para trabajar con señales de audio, en tareas tales como preprocesamiento de voz, grabación de audio, y manipulación de archivos de transcripción (label files). Es importante tener en cuenta que HTK puede también ser usada como una librería de desarrollo.

HTK fue desarrollado en el Departamento de Ingeniería de Cambridge, Cambridge University Engineering Department (CUED). Aunque la principal fortaleza de HTK, es el desarrollo de herramientas para el reconocimiento automático del habla (RAH) y del modelamiento del lenguaje, se puede emplear en las áreas en donde los Modelos Ocultos de Markov sean aplicables, tales como síntesis de voz, reconocimiento de patrones y

estudios de genética.

Debido a las limitantes de tiempo y recursos, no se pudo realizar un entrenamiento intensivo, pues se requeriría de la grabación de cada una de las palabras por miles de personas distintas. En la práctica y entrenamiento del reconocedor, se seleccionaron 50 palabras y 10 voces distintas, 5 masculinas y 5 femeninas, con lo que se obtuvo un reconocimiento aceptable, siempre mayor al 90%.

El código del reconocedor y los scripts para automatizar la reestimación de parámetros en el proceso de entrenamiento y reconocimiento pueden ser encontrados en el Anexo B(Digital).

A continuación se explica como se realizo el reconocedor con la ayuda de HTK

4.4.2.1 Preparación del reconocedor

4.4.2.1.1 Creación de la gramática

El primer paso en la construcción del reconocedor, es la construcción de la gramática, en ella se definen y se acotan las posibles frases y reglas gramaticales de nuestro reconocedor, como ejemplo, una parte de la gramática del reconocedor será:

```
/*definicion de digitos*/
```

```
$digito = [uno] | [dos] | [tres] | [cuatro] | [cinco] | [seis] | [siete] | [ocho] | [nueve] | [cero] ;
```

```
/*Nombres de Amigos*/
```

```
$nombre = [jaime] | [carolina] | [paola] | [diego] | [eivar] | [alex] | [leidy] ;
```

```
/*Definicion de frases para el reconocedor*/
```

```
(sent-start(correo | marca $digito | (telefono | llamar) $nombre) sent-end)
```

4.4.2.1.2 Creación de la red de palabras y el diccionario a partir de la gramática

A partir de la gramática definida se crea la red de palabras posibles a partir del cual se define el diccionario de la aplicación. A continuación se observa parte de las palabras obtenidas y su equivalente fonético después de ser comparadas con las reglas fonéticas del idioma español.

<i>alex</i>	<i>ales</i>
<i>carolina</i>	<i>karolina</i>
<i>cero</i>	<i>sero</i>
<i>cinco</i>	<i>sinko</i>
<i>cuatro</i>	<i>kuatro</i>
<i>diego</i>	<i>diego</i>
<i>dos</i>	<i>dos</i>
<i>eivar</i>	<i>eibar</i>
<i>jaime</i>	<i>jaime</i>
<i>leidy</i>	<i>leidi</i>
<i>llamar</i>	<i>yamar</i>
<i>marca</i>	<i>marka</i>
<i>nueve</i>	<i>nuebe</i>
<i>ocho</i>	<i>oCo</i>
<i>paola</i>	<i>paola</i>
<i>radio</i>	<i>Radio</i>
<i>seis</i>	<i>seis</i>
<i>siete</i>	<i>siete</i>
<i>telefono</i>	<i>telefono</i>
<i>tres</i>	<i>tres</i>
<i>uno</i>	<i>uno</i>

4.4.2.1.3 Creación de Monófonos

A partir del diccionario creado se obtienen el conjunto de monófonos de la aplicación que en el caso serían:

<i>a</i>	<i>r</i>	<i>t</i>	<i>p</i>
<i>l</i>	<i>o</i>	<i>d</i>	<i>R</i>
<i>e</i>	<i>i</i>	<i>g</i>	<i>f</i>
<i>s</i>	<i>n</i>	<i>b</i>	<i>m</i>
<i>k</i>	<i>u</i>	<i>j</i>	<i>y</i>
			<i>C</i>

4.4.2.1.4 Obtención de las grabaciones de entrenamiento

Usando el diccionario, la red de palabras y gramática, se obtiene un listado de algunas frases gramaticalmente validas para realizar el entrenamiento como las que se muestran a continuación:

1. *telefono carolina*
2. *telefono leidy*
3. *telefono carolina*
4. *marca cuatro*
5. *llamar alex*
6. *llamar jaime*
7. *telefono eivar*
8. *marca ocho*
9. *marca tres*
10. *llamar paola*

Ahora se procede a realizar la grabación de las frases, para el trabajo actual se generó un conjunto de 120 frases, de las cuales se usaron 100 para entrenamiento y 20 para la fase de validación.

4.4.2.1.5 Generación de archivos de transcripción

Para que el reconocedor pueda realizar el entrenamiento, debe saber cada archivo de audio a que frase corresponde y así entrenar un modelo oculto para cada fonema. Para tal efecto se construyen archivos de transcripción de la siguiente forma:

```
:"*/1.lab"  
telefono  
carolina  
:"*/2.lab"  
telefono  
leidy  
:"*/3.lab"  
telefono  
carolina  
:"*/4.lab"
```

marca
cuatro
.//5.lab"*
llamar
alex

A partir de este archivo se genera una transcripción fonética para cada una de las grabaciones.

"/1.lab"
sil
t
e
l
e
f
o
n
o
k
a
r
o
l
i
n
a
sil
.

Donde sil significa silencio

4.4.2.1.6 Parametrización de las señales de audio

Finalmente, se procede a la extracción de los vectores de características. Para este caso y debido a la efectividad encontrada en comparación con otros métodos, se decidió usar MFCC o Mel Frequency Cepstral Coefficients que son explicados en el capítulo dos, los cuales basan su funcionamiento en análisis de Fourier. A pesar de que en el recorrido del presente trabajo se desarrolló un módulo extractor de coeficientes cepstrales y se demostró su eficacia, por simplicidad, se decidió usar una herramienta incluida con HTK que permite realizar el mismo procedimiento, llamada Hcopy, que genera los coeficientes

cepstrales para cada archivo de audio.

4.4.2.2 Creación de modelos ocultos de Markov

Una vez se tienen todos los insumos, y se han extraído las características de cada una de las grabaciones de entrenamiento, se procede a generar los modelos ocultos inicializados con datos globales y posteriormente refinados mediante reestimaciones acordes a los datos de entrenamiento hasta obtener el modelo definitivo incluyendo el modelo del silencio. Por último se procede a realizar el reconocimiento buscando la transcripción fonética mas acorde a la muestra tomada.

4.4.2.2.1 Generación de HMM

Debido a que se trabaja con reconocimiento de voz usando fonemas, lo más adecuado es usar un modelo de 3 estados. Para iniciar se toma un modelo base que se ira reestimando cuando se usen los insumos obtenidos para entrenamiento.

A continuación se observa el modelo inicial, que cuenta con 3 estados (2,3,4) y el estado 1 de entrada y el 5 de salida. El tamaño de cada vector es 39, donde se almacenaran los 39 coeficientes cepstrales (13 estáticos, 13 desplazamiento, 13 aceleración). Adicionalmente inicializamos las medias en 0 y las varianzas en 1.

A continuación se puede observar el modelo inicial.

```
~o <VecSize> 39 <MFCC_0_D_A>
~h "prototipo"
<BeginHMM>
<NumStates> 5
<State> 2
  <Mean> 39
  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
  <Variance> 39
  1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
<State> 3
  <Mean> 39
  0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
  <Variance> 39
  1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
```

```

1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
<State> 4
  <Mean> 39
    0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
  <Variance> 39
    1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
<TransP> 5
0.0 1.0 0.0 0.0 0.0
0.0 0.6 0.4 0.0 0.0
0.0 0.0 0.6 0.4 0.0
0.0 0.0 0.0 0.7 0.3
0.0 0.0 0.0 0.0 0.0
<EndHMM>

```

Ahora se procede a hacer la reestimación, usando los coeficientes de las grabaciones realizadas en la fase anterior con la herramienta Hcomp del HTK.

En este momento se cuenta con los modelos ocultos listos, pero todavía no operacionales, pues falta incluirles el modelo de silencio para que nos sean útiles en el reconocimiento.

Una vez se cuenta con los modelos, es necesario realinearlos para posibilitar varias transcripciones por palabra y aumentar la tasa de efectividad. Los detalles, scripts y código de este proceso se pueden observar en el ANEXO B (Digital)

Mejorando a Trifonos en lugar de monófonos

Hasta este punto ya se tienen los modelos reestimados, con el modelo de silencio incluidos, entrenados y listos para ser usados, sin embargo, pueden ser mejorados, usando un modelo probabilístico de trifonos en lugar de monófonos mejorando notablemente su efectividad, por ejemplo:

Palabra	Monófonos	Trifonos
llamar	sil y a m a r	Sil y+a y-a+m a-m+a m-a+r a-r

Usando el modelo de trifonos, la efectividad se aumenta notablemente, pues es sabido

que un fonema difiere mucho de otro, dependiendo del entorno en donde se pronuncia o en otras palabras, dependiendo de la palabra en la que se incluye el fonema.

Es importante anotar, que el uso de trífonos en lugar de monófonos, aumenta notablemente la complejidad y el número de datos a entrenar. En el reconocedor que venimos trabajando, encontramos 21 modelos a entrenar, sin embargo al pasar a trífonos, se podrían alcanzar hasta $21^3=9261$, lógicamente esa sería el máximo de modelos a entrenar, pero acotando el conjunto a los eventos posibles según el diccionario obtenido reduciríamos notablemente el número, para el caso en particular, se obtienen 205.

Se observa así, como el paso de monófonos a trífonos en el caso puntual de este reconocedor, aumenta la complejidad del entrenamiento por un factor de 10, llevando a una carga computacional mayor, que sin embargo será recompensado en la efectividad del reconocedor.

El proceso de llevar de monófonos a trífonos puede ser observado en el ANEXO B(Digital)

4.4.3 Evaluación de resultados

Para la evaluación del reconocedor, se usaron las 20 grabaciones restantes que no fueron usadas para efectos de entrenamiento. Las 20 grabaciones son procesadas y sus vectores de descripción extraídos tal como se realizó con los datos de entrenamiento, después de esto y con ayuda de Hvitte, se realiza la comparación buscando cuáles modelos se acercan más a la representación de cada prueba y se clasifica la frase introducida, realizando su reconocimiento. La aplicación Hresults de HTK nos genera estadísticas de reconocimiento a partir de un archivo que contiene las frases reales que contiene cada grabación de prueba.

Los resultados arrojados fueron de 20 frases introducidas, 19 reconocidas correctamente y 1 incorrectamente para una efectividad de 95%, lo cual supera el requerimiento inicial de 90%.

4.5 SELECCIÓN DE LAS MEJORES CONDICIONES Y PARÁMETROS PARA EFECTOS DE UN ÓPTIMO RECONOCIMIENTO DE LA SEÑAL DE AUDIO

En esta sección se busca establecer una relación directa entre los diferentes parámetros en la red de transmisión y de procesamiento de la señal que redundan en calidad de la voz percibida y el índice de reconocimiento. Para tal efecto, se creó una interfaz entre el sistema de VoWLAN construido en la sección 4.1 y el sistema de reconocimiento construido en la sección 4.4. Se tuvieron en cuenta parámetros como codificación, esquema de paquetización, pérdida de paquetes, retardos, entre otros y se evaluó sobre un conjunto de 200 señales de audio grabadas a través del sistema de VoWLAN para efectos de reconocimiento. Como una referencia, se tomó la tasa de reconocimiento sobre las mismas 200 muestras pero tomadas directamente del micrófono obteniendo un 97.5% de efectividad, provenientes de 195 frases reconocidas correctamente y 5 reconocidas incorrectamente. Es de anotar que esta tasa puede ser mucho mayor, pero necesita de un entrenamiento más grande.

4.5.1 Evaluación de reconocimiento de voz con bajas pérdidas con diferentes codificaciones

Se realizó la evaluación de reconocimiento sin inducir pérdidas adicionales, usando un canal con poca interferencia en donde se presentaron pérdidas siempre inferiores al 1%, incluso cercanas a 0. Para la evaluación se usaron los códecs trabajados a lo largo del documento pero se añadió el códec iLBC usado ampliamente en internet en aplicaciones como skype y google talk, pues según la literatura, este códec comprime cada trama independientemente de las otras, lo cual lo hace, en teoría, más tolerante a pérdidas casuales de paquetes. Los resultados obtenidos se pueden observar en la tabla 4.8 y la figura 4.26.

Códec	G.711	G.729	G.726	G.723(5.3)	G.723(6.3)	iLBC
Tasa de Reconocimiento %	96,5	96,5	95	94,5	96	96,5

Tabla 4.8 Tasa de reconocimiento en canal sin pérdidas por códec

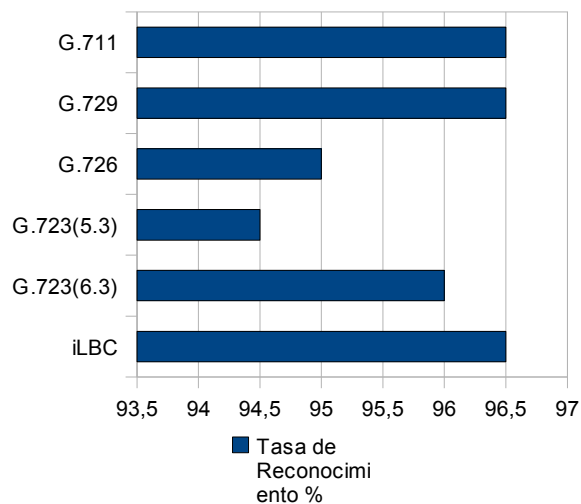


Figura 4.26 Comparativa de tasa de reconocimiento por códec

Se puede observar cómo la tasa de reconocimiento es alta en todos los códecs en un ambiente con pérdidas cercanas a cero. A pesar de esto se puede ver que el uso de g.729 g.711 e iLBC es más adecuado.

4.5.2 Evaluación de reconocimiento de voz con pérdidas con diferentes codificaciones

Con esta evaluación, se busca acercar el sistema a un ambiente más real, induciendo pérdidas de 5, 10 y 15% usando algoritmos PLC para el relleno de muestras perdidas. La inducción de pérdidas se realizó de un modo totalmente aleatorio, donde existe la posibilidad de pérdidas de paquetes adyacentes.

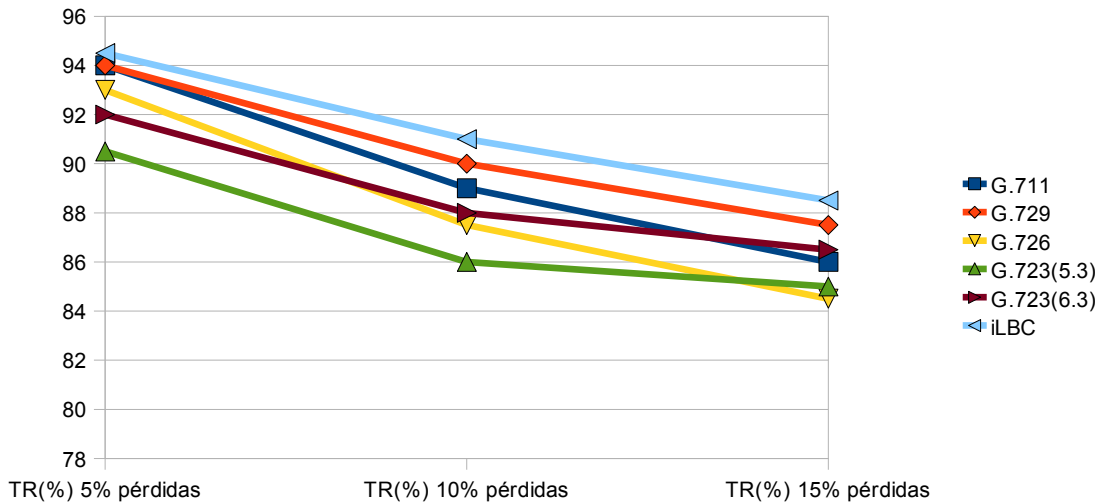


Figura 4.27 Reconocimiento según codificación y cantidad de pérdidas

En la figura 4.27 se puede observar claramente que la tasa de reconocimiento disminuye ante la pérdida de paquetes a pesar de los algoritmos PLC usados. Ante pérdidas cercanas a 0, la tasa de reconocimiento es la misma en g.711 g.729 e iLBC., pero a medida que aumenta la tasa de pérdidas se observa que iLBC es el más adecuado, seguido de cerca por g.729. Se desalienta a usar codificaciones como g.723 en cualquier caso o g.711 cuando las pérdidas son altas.

Es importante aclarar que lo que se busca medir es la variación de la tasa de reconocimiento ante la variación del porcentaje de pérdidas y de codificación. La tasa de reconocimiento obteniendo la voz directamente de la fuente es solo un referente, de modo que si se realiza un mayor entrenamiento del sistema y la tasa de reconocimiento aumenta, también lo deben hacer con las señales codificadas y con pérdidas.

5 ESQUEMA DE IMPLEMENTACIÓN DE SISTEMAS DE RECONOCIMIENTO DE VOZ SOBRE VOWLAN PARA EL IDIOMA ESPAÑOL

A partir de los resultados de los trabajos expuestos y los desarrollos y montajes realizados, se concluye en este capítulo con la propuesta del esquema objeto de esta tesis, el posterior montaje del prototipo correspondiente y la evaluación de los resultados obtenidos que validan el modelo propuesto.

5.1 PROPUESTA DE ESQUEMA DE OPERACIÓN PARA SOLUCIÓN DE RECONOCIMIENTO DE VOZ OPERANDO SOBRE VOWLAN

El esquema propuesto abarca la adaptación y ajuste de parámetros a diferentes niveles en cada uno de los módulos que comprenden la operación de un sistema de reconocimiento de voz sobre una red de VoWLAN. Como propuesta, se dividió el sistema en 5 bloques funcionales principales que se muestran en la figura 5.1. A continuación se detalla la función de cada uno de los bloques

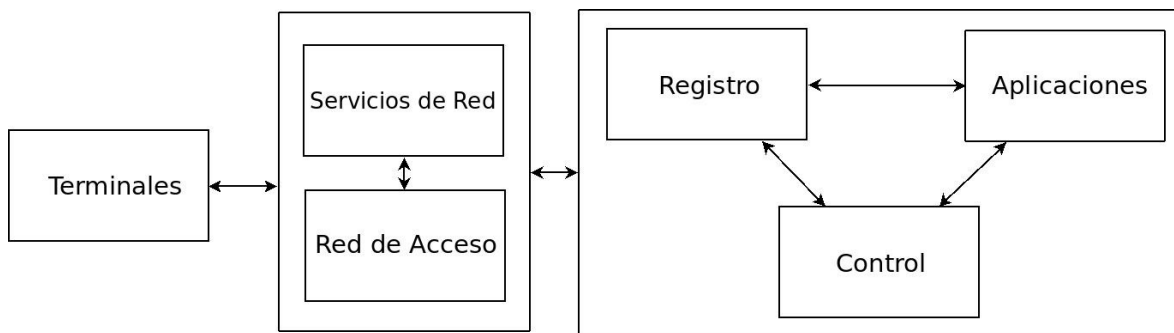


Figura 5.1 Esquema modular sistema de vowlan

5.1.1 Módulo Registro

El módulo de registro se encarga de obtener la información de los terminales que desean acceder al sistema, sus funciones son:

5.1.1.1 Validación del cliente

La validación del cliente se realiza a través de un servicio de validación como Radius o LDAP, esto brinda ventajas referentes a la gestión del sistema y generación de reportes de uso e intentos de acceso no autorizados, además de proveer de datos estadísticos útiles para la futura depuración del sistema.

5.1.1.2 Información de estado

Este módulo podrá reportar el estado de cualquier cliente en cualquier momento, y podrá negociar nuevas condiciones para el cliente deseado, como cambio de codificación o de estado. Estos cambios podrán ser realizados por el módulo de control.

5.1.1.3 Almacenamiento de información de registro incluyendo:

El módulo también almacena información concerniente a los clientes y sus datos de registro, debe contener al menos los siguientes:

1. Identificador o login del usuario
2. Dirección IP de registro
3. Tipo de Conexión (Directa/NAT)
4. Códecs permitidos por el cliente

Esta información será usada por el módulo de aplicación y por el módulo de control.

5.1.2 Módulo Control

Este módulo implementa una extensión de funcionalidades al sistema de comunicaciones IP tradicional e interactúa directamente con el módulo de aplicaciones y el módulo de registro. Está compuesto por dos submódulos y un repositorio o base de datos con información referente al estado de los terminales que solicitan servicios al servidor de aplicaciones.

5.1.2.1 Submódulo de identificación de parámetros de conexión

Este submódulo monitorea el estado y condiciones de los clientes que solicitan la ejecución de una aplicación al módulo de aplicaciones y los registra en la base de datos de estados de conexión. Este submódulo puede también bajo demanda del módulo de aplicaciones verificar el estado de conexión de cualquiera de los clientes registrado, obtiene y almacena en la base de datos como mínimo los siguientes datos de cada cliente:

- Delay
- Jitter
- Pérdidas
- Punto de acceso asociado.

5.1.2.2 Submódulo de Control de Capacidad

Este módulo toma decisiones basándose en los datos almacenados en la base de datos de estados de conexión, a partir de éstos toma una de las siguientes decisiones.

- **Permitir la llamada:**

Ante unas buenas condiciones de enlace, delay, jitter y bajas pérdidas y bajo uso del canal inalámbrico, se permite la llamada sin restricción alguna y los terminales negocian el códec y condiciones de la llamada.

- **Permitir la llamada, restringiendo parámetros de conexión**

Ante condiciones no óptimas como altos delays, jitter o pérdidas, se mitigará el efecto tomando decisiones según una base de reglas construida a partir los

resultados obtenidos en el capítulo 4. Las limitaciones se harán modificando temporalmente el registro de los terminales, restringiendo los códecs a usar, modificando el jitter buffer o activando mecanismos de PLC

- **No permitir la llamada.**

Si las condiciones son demasiado adversas o se ha alcanzado el límite de capacidad por AP wifi, se da aviso al terminal y se rechaza la llamada.

5.1.2.3 Base de datos de estados de conexión

Esta base de datos contiene la información proveniente del submódulo de identificación de parámetros de conexión. Adicionalmente mantiene un histórico y puede ser usada para la identificación de problemas y generación de reportes. Es usada como insumo principal para la toma de decisiones del submódulo de control de capacidad.

5.1.3 Módulo de Aplicaciones

El módulo de aplicaciones es llamado bajo demanda de los clientes ante la petición de un servicio como la solicitud de una llamada de un terminal a otro o la consulta del correo de voz. Como primera instancia, éste módulo se comunica con el módulo de control el cual determina la ejecución o no de la aplicación solicitada. Este módulo controla la interacción de cada uno de los clientes con otros o con servicios implementados en aplicaciones de correo de voz, conferencia, reconocimiento y síntesis de voz, entre otros.

5.1.4 Módulo de Servicios de Red

Para el óptimo funcionamiento del sistema, es necesario proveer servicios de red que permitan la comunicación entre los diferentes elementos del sistema como equipos clientes, puntos de acceso inalámbricos, servidores de registro, servidor de aplicaciones, entre otros. Aunque existe una gran cantidad de servicios de red que pueden ser implementados para aumentar las funcionalidades del sistema, tres de ellos son imprescindibles:

5.1.4.1 DHCP

Es necesario implementar un sistema de asignación dinámico de direcciones y parámetros de red, de modo que cada cliente obtenga su configuración de red de modo automático y pueda acceder al sistema. Es necesario un servidor centralizado dedicado a

este servicio, pues si se delega la función a cada uno de los puntos de acceso, no será posible el roaming, ya que los parámetros de conexión cambiarían dependiendo del AP al que se asociara, provocando desconexiones una vez se exceda el TTL de la delegación DHCP.

5.1.4.2 AAA

Se debe implantar un servicio de AAA(Authentication, Authorizing and Accounting) que controle y autorice dos elementos:

5.1.4.2.1 Acceso a la red inalámbrica

Para aumentar en nivel de seguridad y encriptación de los datos en la red inalámbrica, se usará un sistema de autenticación basado en WPA o WPA2 con 802.1X, el cual requiere de un servidor AAA para validar parejas de login/password. A través del AAA se podrán validar atributos adicionales correspondientes a cada cuenta de usuario, como estado de la cuenta, expiración, permisos, entre otros.

5.1.4.2.2 Autenticación en el servidor de registro

Una vez el cliente es autenticado e ingresado a la red inalámbrica, este procederá a validarse en el sistema de registro quien a su vez, validará el acceso a través del AAA

5.1.4.3 DNS

Los servidores a los que se desea ingresar son llamados por nombres, por lo cual es necesario el uso de servicio DNS para traducir los nombres a direcciones IP y para permitir también el uso de cluster para alta disponibilidad o para alto desempeño al registrar varias direcciones IP a un nombre y administrar las consultas con mecanismos round robin. Para servicios especiales como de voz ip, se crean registros especiales de tipo srv.

5.1.5 Red de Acceso

El módulo de red de acceso comprende todos los elementos de red que permiten la comunicación entre los terminales 802.11 y los equipos y servidores de comunicaciones.

Su principal objetivo es el de garantizar una conexión fiable y que cumpla con los requerimientos para poder transmitir VoIP sobre el. Los principales parámetros a tener en

cuenta en la implementación de este módulo son:

5.1.5.1 Cobertura

Para tener una cobertura amplia del sistema, se debe dotar al sistema de equipos de comunicaciones inalámbricas adecuados, con arreglos de antenas con una polarización adecuada a la zona que se busca darle cobertura y con una potencia de radiación que no sobrepase las establecidas por las regulaciones vigentes. El tema de cobertura está directamente relacionado con el de capacidad, pues aunque se puede dar una amplia cobertura con un solo punto de acceso, amplificadores, repetidores y antenas de altas ganancias, el echo de pertenecer al mismo canal y SSID, limita la capacidad en servicios de VoIP a alrededor de una docena de llamadas, que pueden ser insuficientes para una zona grande de cubrimiento. Para poder lograr una cobertura amplia y sin tener una limitante tan grande de capacidad, se pueden usar celdas de cobertura, cada una operando en un canal distinto, buscando siempre que las celdas adyacentes estén espaciadas 3 o por lo menos 2 canales. Como la idea es ampliar la cobertura, en el diseño se determinan puntos frontera para cada celda, cuidando minimizar la cantidad de zonas muertas y proveyendo de zonas de translape entre los canales donde los clientes llevan a cabo el roaming. Aunque de esta forma se soluciona el inconveniente de capacidad, se introduce un nuevo factor a tener en cuenta que es el retardo adicional ocasionado por roaming. Este retardo surge en el momento en que un cliente cambia su asociación de un canal a otro por encontrar un mejor nivel de señal en el segundo. Este tiempo normalmente es de menos de 100ms.

5.1.5.2 Capacidad

El uso de varios puntos de acceso usando celdas que se translapan entre si, permite un aumento de la capacidad pues cada celda maneja su límite propio, de modo que se podrían tener alrededor de 12-15 conversaciones concurrentes por cada celda. Si se requiere ampliar la cantidad de llamadas por zona, bastará con aumentar el número de celdas y hacer cada una de ellas más pequeña. En casos donde la densidad de llamadas sea demasiado alta, se pueden poner hasta 3 celdas ubicadas en la misma zona física, con espaciamento de 3 canales entre ellas para evitar problemas de interferencia, permitiendo al rededor de 45 llamadas concurrentes por zona.

5.1.5.3 Calidad del Servicio

Debido a que el esquema debe ser implementable en cualquier tipo de red, no se puede limitar la operación a la existencia de mecanismos de control de tráfico o marcado de paquetes que realizan funciones específicas de calidad de servicio, así que se limitará a proveer las mejores condiciones para que el tráfico de voz sea llevado a su destino de la mejor forma posible, evitando o minimizando fenómenos como pérdidas, retardos y jitter altos.

5.1.5.3.1 Modo de Operación

El modo de operación será el modo G o modo B, pero en ningún momento, modo mixto. El uso de modo mixto, disminuye la capacidad de llamadas del sistema en más del 25% (según pruebas realizadas en laboratorio) con el único beneficio de permitir clientes 802.11b los cuales están haciéndose obsoletos rápidamente. Si se quiere dar soporte a equipos que solo soportan 802.11b como dispositivos móviles o teléfonos wifi antiguos, se recomienda el uso de puntos de acceso con dos radios para que manejen por separado los dos modos.

5.1.5.3.2 Interferencias

El diseño del sistema inalámbrico debe buscar proveer siempre la mejor relación señal a ruido posible, evitando así, pérdidas altas de paquetes y la consecuente degradación de la calidad de la comunicación. Se deben buscar arreglos de celdas separadas varios canales entre sí. Si es necesario la comunicación entre puntos de acceso por un medio inalámbrico, se debe usar comunicación por otra banda, en el caso wifi, se recomienda 802.11a en la banda de los 5GHz.

5.1.5.3.3 Mínimas condiciones de asociación

Para proveer una calidad de servicio adecuada, se delimita la operación del sistema de la siguiente forma:

Velocidad mínima de asociación 11Mbps, con una velocidad menor, además de disminuir la capacidad, se es más sensible a interferencias ya que es requerida una relación señal a ruido mucho menor para asociarse a 1Mbps que para asociarse a 11 o 12Mbps

Autenticación y Encriptación: WPA/WPA2 802.1X, debido a las altas fallencias en el protocolo WEP y a la dificultad de negociación de llaves o certificados digitales, se opta

por usar un mecanismo basado en WPA/WPA2 que use 802.1X permitiendo la validación login/password a través de servidores AAA externos.

5.1.6 Terminales

Cualquier equipo terminal que soporte 802.11b/g y pueda trabajar como un cliente de VoIP podrá ser cliente del sistema. Debido a que el esquema planteado debe operar sobre redes, servidores y clientes ya existentes, no se puede limitar el tipo de clientes a usar, sin embargo, como trabajo opcional del módulo se puede proponer como alternativa la operación distribuida entre terminales, equipos de red y servidores para labores como ajuste de parámetros de conexión, control de capacidad e incluso extracción de características de la señal de voz en el terminal y el envío de los vectores como datos a la aplicación de reconocimiento.

5.2 PROTOTIPO DE UN SISTEMA DE RECONOCIMIENTO DE VOZ SOBRE VOWLAN

Usando los trabajos realizados en el capítulo 4, se procedió a conformar un sistema completo de VoIP sobre WLAN, totalmente acorde al esquema y recomendaciones propuestas en la sección anterior. La figura X muestra los módulos que conforman el sistema.

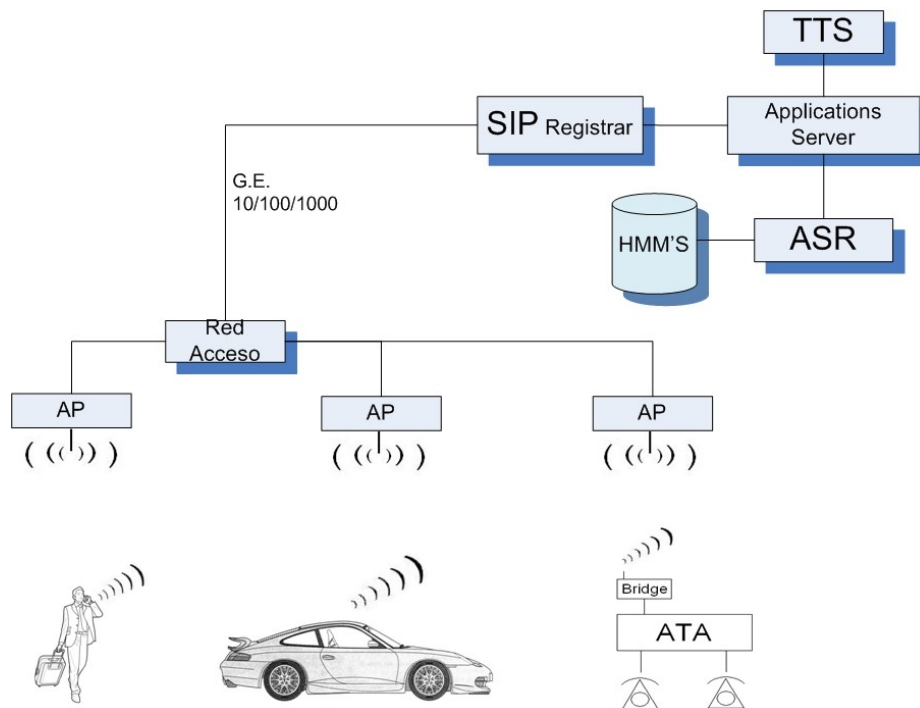


Figura 5.1 Diagrama general del prototipo

A continuación se desglosa cada uno de los módulos y se explica el funcionamiento de los mismos.

5.2.1 SIP Registrar

Este módulo inicialmente estuvo conformado por el módulo de registro de la PBX Asterisk, de código abierto, sin embargo, para integrarlo al prototipo y que pudiera cumplir con las condiciones mínimas de calidad de la comunicación y de la señal de audio recibida y luego ser procesada correctamente por el módulo de reconocimiento, fueron creados dos módulos adicionales: Control de Capacidad y Medición de parámetros de conexión.

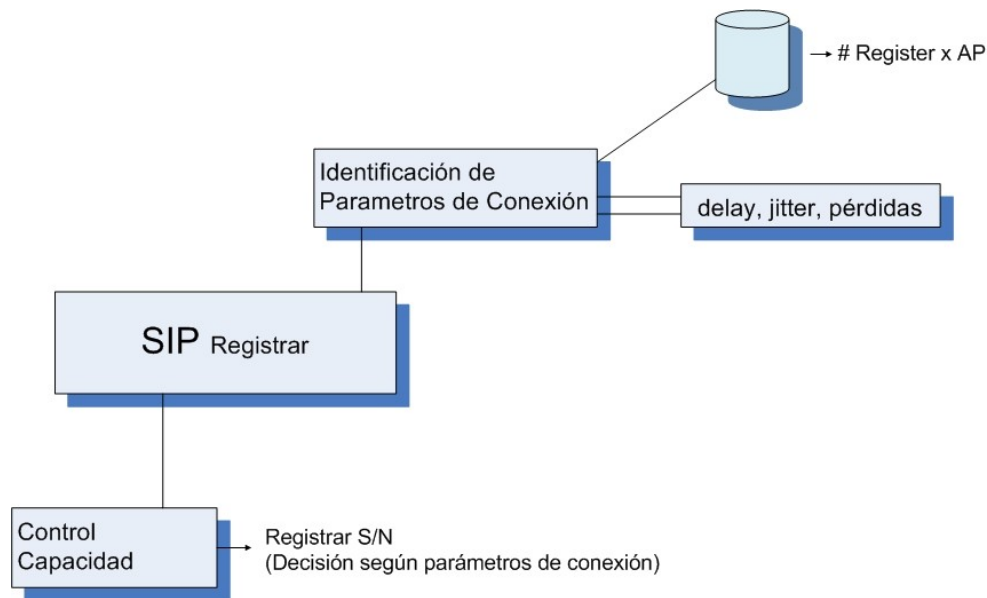


Figura 5.2 Implementación del módulo de registro y nuevos submódulos

5.2.1.1 Módulo de registro simple

Este módulo es proveído por la PBX Asterisk, es encargado de realizar entre otros procesos, el registro de los clientes, incluyendo datos como dirección IP, login y códecs disponibles y por lo tanto, permite la ubicación de los clientes cuando se realiza una llamada hacia ellos. Este módulo está conectado al servidor de aplicaciones y al módulo de control de capacidad. Funciona como cualquier sip registrar, pero antes de permitir el registro del terminal consulta un bit generado por el control de capacidad, en el cual se determina si la llamada puede ser cursada o no.

5.2.1.2 Módulo de medición de parámetros de calidad

En este módulo se obtienen y almacenan los valores referentes a delay, jitter y pérdidas de los participantes en la llamada que se está solicitando. Así mismo, se identifica a que punto de acceso inalámbrico está asociado cada uno de los clientes y cuales de ellos están cursando llamadas en el momento. El módulo se encarga de mantener actualizada la base de datos de parámetros la cual es consultada periódicamente por el control de capacidad. Para la toma de estas medidas, se escribió una aplicación en Python que

obtiene parámetros de la calidad de la conexión como la cantidad de paquetes perdidos y usa el Asterisk Manager API para la obtención de parámetros específicos de llamadas en curso como delay y jitter. El código de la aplicación puede observarse en el Anexo B (Digital).

5.2.1.2.1 Base de Datos de parámetros de calidad

En esta base de datos se almacenan los datos proveídos por el módulo de medición de parámetros de calidad, estos datos son consultados posteriormente por el módulo de control de capacidad. La base de datos fue implementada usando MySQL 5.0, la estructura de la base de datos puede observarse en el Anexo B (Digital).

5.2.1.3 Módulo de control de capacidad

Este módulo se encarga de revisar las condiciones de los clientes previo a la generación de la llamada. Usa como insumos los datos almacenados en la base de datos de parámetros de calidad una base de reglas para la toma de decisiones. El módulo puede permitir la llamada directamente, permitir la llamada después de cambios en datos de registro o denegar la llamada.

El detalle de las condiciones referentes a la toma de decisiones y el código realizado en python puede encontrarse en el Anexo B (Digital)

5.2.2 Servidor de Aplicaciones

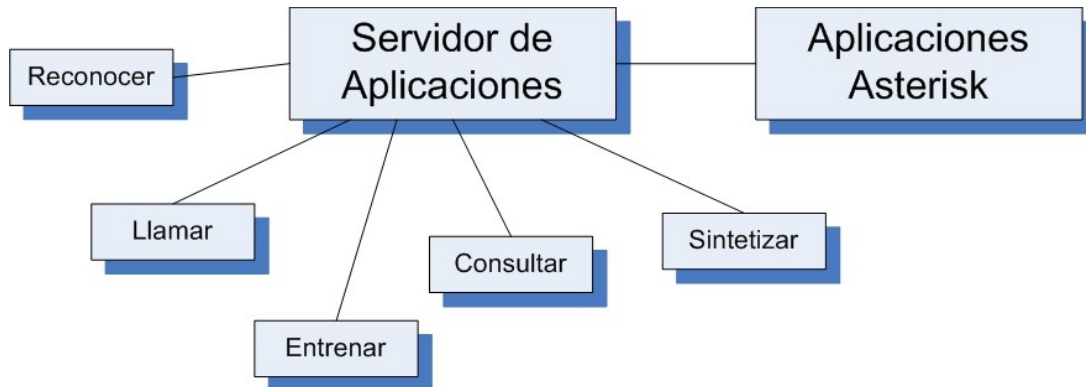


Figura 5.3 Servidor de aplicaciones y aplicaciones desarrolladas para el prototipo

En la figura 5.3 se pueden observar las principales aplicaciones implementadas en el prototipo. El servidor de aplicaciones está conectado directamente al servidor de registro. Para los procesos típicos como realizar una llamada, correo de voz, conferencia, entre otras, se usan las aplicaciones ya incluidas en Asterisk, sin embargo se introdujo como condición previa a las aplicaciones la solicitud de aprobación del módulo de control de capacidad.

Las aplicaciones escritas para la operación del prototipo son las siguientes:

5.2.2.1 Entrenar:

Aplicación diseñada para facilitar el entrenamiento, hace uso de la aplicación sintetizar para guiar al usuario en el entrenamiento indicándole las frases a pronunciar. Usa las aplicaciones y scripts de la sección 4.4 destinadas a entrenamiento para alimentar la base de datos de HMM's

5.2.2.2 Reconocer:

Esta aplicación captura la señal de audio permanentemente y la procesa con la aplicación de reconocimiento. Una vez reconocida la frase la procesa y ejecuta la aplicación que corresponda como "consultar jaime" o "llamar carolina". Usa las aplicaciones y scripts de

la sección 4.4 destinadas a reconocimiento.

5.2.2.3 Llamar:

Esta aplicación se ejecuta una vez la aplicación reconocer detecta la palabra llamar y el destino de la llamada, es una aplicación muy sencilla pues usa la aplicación Dial de asterisk para su funcionamiento

5.2.2.4 Consultar:

Una vez detectado el llamado en la aplicación reconocer, busca en la base de datos el número telefónico correspondiente al usuario consultado.

5.2.2.5 Sintetizar:

Esta aplicación se encarga de generar voz de respuesta a las diferentes aplicaciones a partir de texto y resultados de base de datos. Usa el Sintetizador Festival para tal efecto y conjunto de voces en español de libre distribución.

5.2.3 ASR

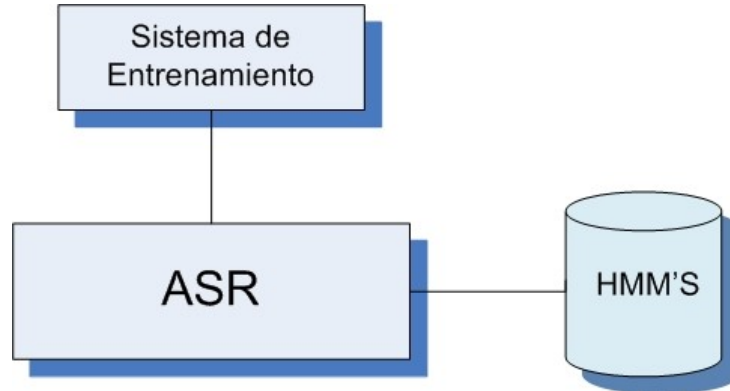


Figura 5.4 Sistema de Reconocimiento del habla

El sistema de reconocimiento creado en el capítulo 4 fue usado en el prototipo(figura 5.4). Se crearon una serie de interfaces para el paso de la información de audio al reconocedor y la entrega de resultados del reconocedor a las frases introducidas. Se creó una aplicación para realizar un entrenamiento sencillo usando la interfaz telefónica. Todos los modelos creados se almacenan mejorando constantemente la efectividad del reconocedor.

5.2.4 TTS

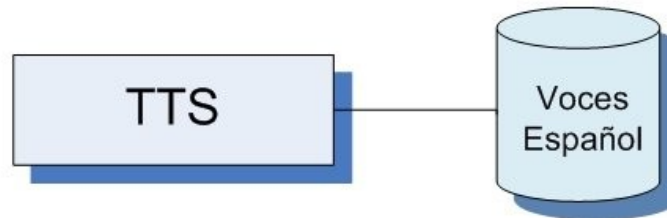


Figura 5.5 Sistema de Síntesis de Voz

Como sintetizador se usó Festival y se usó una aplicación existente de libre de distribución que captura la señal de audio sintetizada y la envía al cliente asterisk.

5.2.5 Acceso inalámbrico

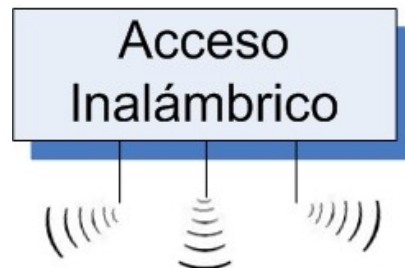


Figura 5.6 Sistema de Acceso Inalámbrico

En cuanto a los puntos de acceso se usó un arreglo de 3 Aps con antenas tipo panel y lóbulo de radiación de 120 grados, en tres configuraciones distintas:

1. Modo Hotspot:

Se cubrió una zona completa de 360 grados, usando los tres puntos de acceso con sus respectivas antenas, los canales usados fueron 1,6 y 11

2. Modo Movilidad:

Se ubicaron los Ap y antenas buscando cubrir un área de circulación, imitando el cubrimiento en carretera o en zonas de tránsito. Se usaron canales 1,6 y 11 pues la primera celda no alcanza a interferir con la tercera.

3. Modo Alta densidad:

Se ubicaron los Ap para proveer la máxima densidad de llamadas posibles, se cubrió una misma zona con los tres puntos de acceso usando canales 1,6 y 11.

Se establecieron las condiciones propuestas en el modelo, estableciendo como frontera una velocidad mínima de asociación 12Mbps y modo de operación solo G.

Los puntos de acceso usados fueron equipos linksys con sistema operativo linux, lo que permitió evaluar adicionalmente el efecto de incorporar mecanismos de calidad de servicio en la red inalámbrica usando el controlador de tráfico TC.

5.3 PRUEBAS AL PROTOTIPO

El prototipo implementado cumple a cabalidad con el modelo propuesto, y fue construido con base en el. Para evaluar el funcionamiento, se tomo como referencia un sistema de VoWLAN típico sin la implementación de los módulos y recomendaciones realizadas en el esquema. A continuación se describe el sistema típico contra el que se evaluó el sistema prototipo.

5.3.1 Descripción del sistema típico

Para la evaluación del sistema prototipo, se debió realizar el montaje de un sistema típico contra el cual se pudiera comparar el sistema prototipo generado a partir del sistema propuesto. De esta forma se dispuso de un sistema de referencia sobre el cual comparar y observar las diferencias en respuestas, tiempos, retardos y otros parámetros que afectan notablemente la tasa de reconocimiento obtenida.

Un sistema típico de red inalámbrica no dispone de mecanismos de calidad de servicio y esta hecho para transmitir datos, por lo cual no presenta condiciones especiales para el transporte de la voz o cualquier otro tráfico de tiempo real. El estándar 802.11b/g no

puede garantizar calidad de servicio debido a que a nivel físico todo el tráfico es tratado de la misma forma y el control que se logre hacer en cuanto a priorización de tráfico a un nivel superior, será insuficiente y no proveerá las condiciones necesarias para garantizar que el tráfico de voz tenga un tránsito adecuado en condiciones de saturación.

El sistema típico carece de un sistema de control, pues está concebido para transmitir datos y entre ellos, la voz como un flujo normal de datos, por lo que no puede limitar el número de llamadas concurrentes ni interactuar con el sistema limitando u optimizando los parámetros de conexión a diferencia del sistema prototipo que si incorpora un mecanismo de control para tal efecto.

El sistema de registro carece de base de datos de parámetros de conexión y de la asociación con un mecanismo de control.

El servidor de aplicaciones es el mismo del sistema prototipo y no presenta ninguna variante.

En la tabla 5.1 se resumen los principales parámetros de conexión del sistema típico y sistema prototipo:

Característica	Sistema Típico	Sistema Prototipo
Estándar wifi	802.11b/g/mixto	802.11g
Velocidades de negociación	54, 48, 36, 24, 18, 12, 11, 9, 6, 5.5, 2, 1Mbps	54, 48, 36, 24, 18, 12
SNR mínima de asociación	15dB	25dB
Códecs	Cualquiera	Escogido por módulo de control
Control de conexiones	No dispone	Módulo de control
Delay máximo	Sin límite	150ms(cada trayecto)
Jitter máximo	Sin límite	50ms (Ida y vuelta)
Autenticación	Cualquiera	WPA/WPA2
ASR	Probabilístico (HMMs)	Probabilístico (HMMs)
Tráfico	En tiempo real (VoIP)	En tiempo real (VoIP)

Tabla 5.1 principales parámetros de conexión del sistema típico y sistema prototipo

5.3.2 Operación en condiciones promedio pérdidas muy bajas

En la tabla W se muestran los resultados en cuanto a tasa de reconocimiento para un sistema típico de VoWLAN y para el prototipo implementado a partir del esquema. Es de anotar que el sistema típico operó en modo mixto, mientras el prototipo se limitó a modo G según recomendaciones del esquema. La tabla 5.1 y figura 5.7 ilustran mejor los resultados

Número de Llamada	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Sistema Típico (%)	96,5	96,5	95,5	96	94	95,5	96	96	72	54,5	16,5	0	0	0	0
Sistema Prototipo (%)	96,5	96,5	96,5	96,5	96	96,5	96	96	95,5	96,5	96	96,5	94,5	NP	NP

Tabla 5.1 Comparación de Tasa de reconocimiento condiciones promedio

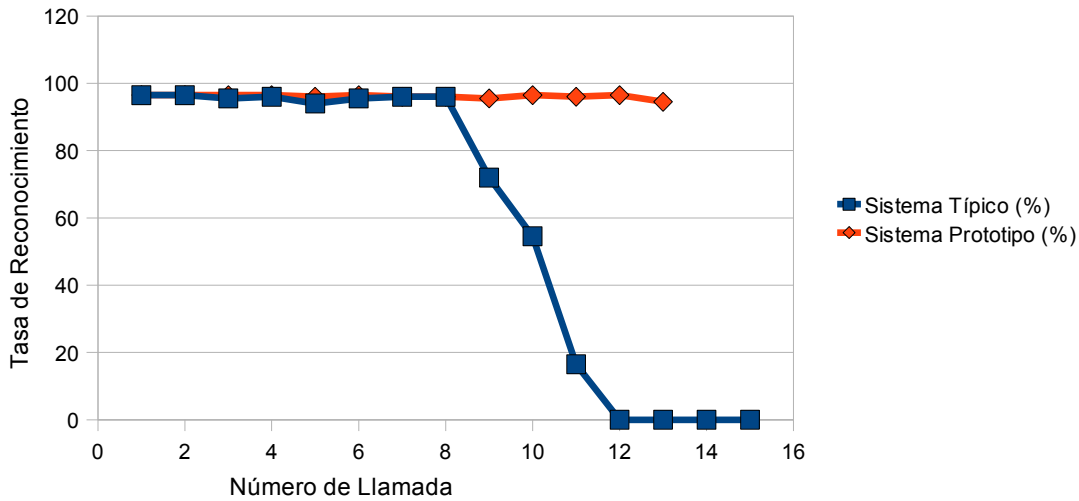


Figura 5.7 Comparación de Tasa de reconocimiento condiciones promedio

Para la obtención de resultados, se probó el sistema con 200 frases, generadas en cada una de las llamadas y se tabularon los resultados obtenidos. Se observa como la tasa de reconocimiento es similar o ligeramente mayor en el sistema prototipo cuando hay pocas llamadas en el sistema, sin embargo, a partir de la llamada 8 las tasas del sistema típico disminuyen abruptamente. En el caso del sistema prototipo, se mantiene una tasa de reconocimiento siempre alta, incluso hasta su última llamada, la número 13. A partir de ahí, el sistema no permite la ejecución de nuevas llamadas.

Para la evaluación se trató de emular una situación típica, con clientes de diferentes tipos y configuraciones.

Son muchos los factores que dan lugar a estos resultados, se pueden resaltar los siguientes:

- El prototipo siempre busca la negociación de un códec adecuado para efectos de reconocimiento, el sistema típico usa el que prefieren los clientes.
- El prototipo no permite la operación de modo mixto, el sistema típico disminuye su capacidad una vez un equipo se asocia en modo 802.11b
- El prototipo establece condiciones mínimas de asociación en cuanto a velocidad, redundando en mejores condiciones señal a ruido. El sistema típico permite asociar clientes bajo cualquier condición
- El prototipo establece un sistema de control para que las llamadas sean cursadas solo cuando se presenten condiciones adecuadas, el sistema típico permite que las llamadas sean cursadas sin establecer condiciones mínimas, lo que repercute en un deterioro de la calidad en todo el sistema.

5.3.3 Operación en condiciones promedio pérdidas muy bajas modo G

Uno de los factores que mayor impacto tuvo en la tasa de reconocimiento promedio en el caso anterior fue la operación en modo mixto. En esta prueba se comparan los reconocedores limitando el sistema típico a la operación en modo solo G.

Número de Llamada	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Sistema Típico Modo G (%)	96,5	96,5	95,5	96	95,5	95	95	95,5	94,5	93	92	79	41	12,5	0
Sistema Prototipo (%)	96,5	96,5	96,5	96,5	96	96,5	96	96	95,5	96,5	96	96,5	94,5	NP	NP

Tabla 5.2 Comparación de Tasa de reconocimiento condiciones promedio, modo G

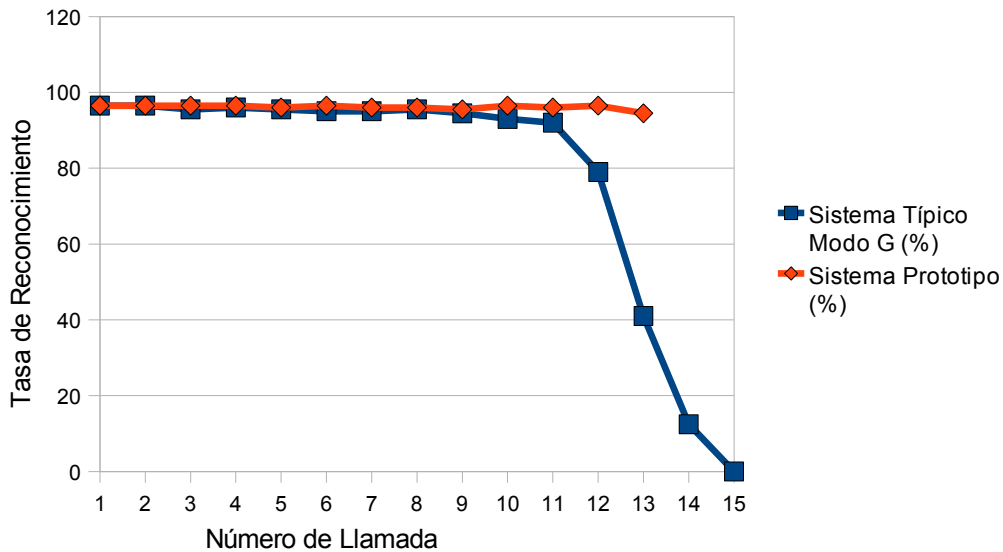


Figura 5.8 Comparación de Tasa de reconocimiento condiciones promedio, modo G

Los resultados se pueden observar en la tabla 5.2 y figura 5.8

5.3.4 Operación en condiciones promedio, con pérdidas debido a interferencias de canales adyacentes

Debido a que el esquema no se limita solamente a la implementación de módulos de control y aplicaciones, sino que también da recomendaciones sobre el diseño e implementación de la red de acceso y servicios, no es posible mostrar de un modo científicamente válido las mejoras que estas recomendaciones puedan representar pues serían totalmente dependientes del sistema específico que se entre a evaluar. Sin embargo, para recrear e ilustrar con un ejemplo se crearon unas condiciones comunes en que opera un sistema VoWLAN y se muestra como adaptando el sistema al esquema se logra una mejora considerable en la tasa de reconocimiento.

Las condiciones del sistema son las siguientes:

- Operación en canal 6, interferencias causadas por redes wifi operando en canales 4 y 5
- Asociación a partir de 1Mbps
- Codificaciones presentes en los clientes G711, G729, G723, iLBC

Se tomaron las tasas de reconocimiento y después se adaptó el sistema al esquema propuesto, haciendo un rediseño del enlace inalámbrico, poniendo la operación en el canal 9 y la asociación mínima a partir de 12Mbps, además de incluir los módulos de control y el esquema en general. Las mejoras se reflejan en la tabla 5.3 y figura 5.9.

Número de Llamada	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Sistema Típico (%)	95	95,5	95,5	94,5	91,5	92	92	89,5	68	43,5	6,5	0	0	0	0
Sistema Prototipo (%)	96,5	96,5	96,5	96	96,5	96	96	96,5	96	96,5	95,5	96	94	NP	NP

Tabla 5.3 Comparación de Tasa de reconocimiento condiciones promedio, pérdidas promedio

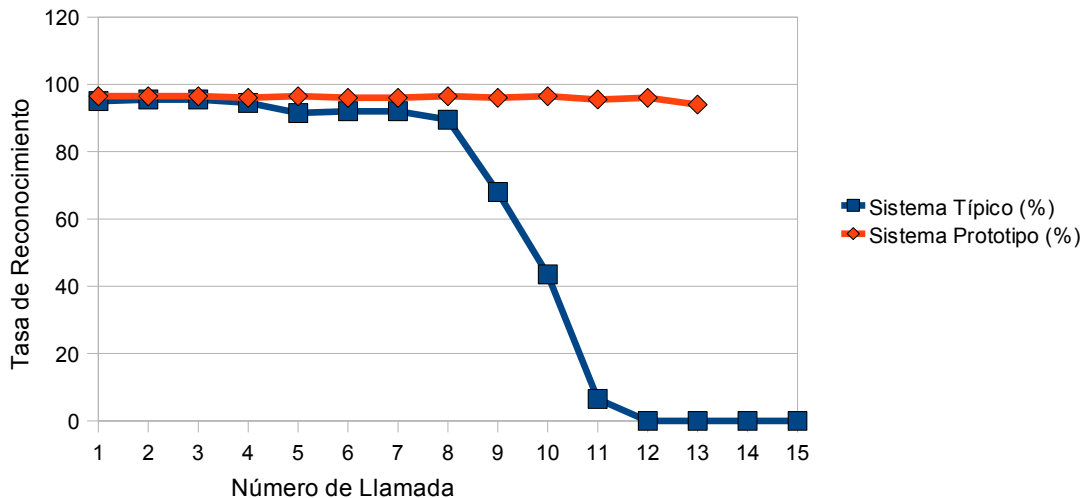


Figura 5.9 Comparación de Tasa de reconocimiento condiciones promedio, pérdidas promedio

5.3.5 Conclusión a resultados de las pruebas

Como se observa en las pruebas y gráficas realizadas, en todos los casos se obtuvo una mejora marcada en la tasa de reconocimiento cuando se adapta el sistema de VoWLAN al esquema propuesto. Incluso en el caso en que se usa modo solo G, las diferencias son apreciables, obteniendo además de varios puntos adicionales en la tasa de reconocimiento, una estabilidad en el funcionamiento del sistema, pues el módulo de control no permite que se sobrepase la capacidad del mismo.

CONCLUSIONES Y FUTURAS LINEAS DE INVESTIGACION

El desarrollo de este proyecto de tesis de maestría ha permitido concluir que:

- El uso de esquemas y medidas de control sobre un sistema VoWLAN garantiza el óptimo funcionamiento del mismo, ofreciendo una alta calidad en el servicio, las comunicaciones de voz y la calidad de audio de la señal lo que además de permitir el uso de aplicaciones de reconocimiento del habla, ofrece una buena percepción al usuario final.
- El factor de mayor incidencia en la tasa de reconocimiento de voz de una señal transmitida a través de un sistema de VoWLAN es la pérdida de paquetes, aún más que el tipo de codificación o jitter existente.
- En un sistema de VoIP que trabaja sobre redes con alta probabilidad de pérdida de paquetes como redes inalámbricas, se debe trabajar con códecs tolerantes a pérdidas y pocas muestras por paquetes, aún a expensas del alto consumo de ancho de banda que esto pueda acarrear.
- Las comunicaciones de VoIP en la redes inalámbricas tienen límites abruptos, esto quiere decir que luego que se supera el límite de capacidad, una comunicación siguiente queda sin ninguna probabilidad de ser soportada a razón del retardo y pérdida de paquetes tan significativo que se presenta, por esta razón, se debe tener mecanismos de control en pro de ofrecer confiabilidad de las comunicaciones y calidad de las señales de voz.
- La construcción de un sistema de reconocimiento del habla puede ser realizada de

muchas formas, sin embargo los métodos probabilísticos como el uso de HMMs, se ajustan adecuadamente al reconocimiento de señales en las que puede haber distorsión o pérdida de fidelidad de la señal, pues el reconocimiento se basa en las transiciones y probabilidad de eventos. Para mitigar el efecto de pérdida de paquetes en una señal de voz y su consecuente descripción incorrecta por parte del descriptor, se deben usar técnicas de PLC que aproximan el contenido de la fracción de audio perdida.

- Se debe buscar a toda costa maximizar la relación señal a ruido y de esta forma disminuir la pérdida de paquetes. Para esto se debe hacer un diseño óptimo de la red inalámbrica ubicando fronteras adecuadas y estableciendo condiciones mínimas de asociación.
- Aunque el prototipo se construyó usando un reconocedor desarrollado y entrenado para operar sobre el idioma español, no fue del alcance del mismo profundizar sobre el idioma de implementación. Se propone como trabajo de investigación futuro, la optimización del sistema de reconocimiento para operar específicamente sobre el idioma español.
- El alcance del proyecto propone un esquema de operación implementable sobre redes WLAN existentes sin ninguna precondición referente al soporte de mecanismos de QoS como marcado de paquetes o diferenciación de servicios. Se plantea como trabajo de investigación, la evaluación y uso de estas características como adición al esquema propuesto con el objetivo de mejorar el desempeño del sistema y garantizar que el tráfico de datos en la misma red no produzca efectos nocivos importantes sobre el tráfico de voz.
- Se concluye que para efectos de reconocimiento del habla, el tratamiento de la señal en todo el sistema, lleva a su degradación y por lo tanto a una disminución en la tasa de reconocimiento. El esquema propuesto minimiza esta degradación para que la señal llegue lo mas fiel posible al sistema reconocedor y pueda ser descrita correctamente por el sistema de descripción. Se propone como línea de investigación el reconocimiento del habla distribuido, donde se extraigan las características de la señal de voz en el terminal y se envíen los vectores de descripción directamente al reconocedor, de modo que los efectos nocivos en el tratamiento de la señal no tengan ningún efecto en la tasa de reconocimiento final.

REFERENCIAS

- [1] L. RABINER and BIING-HWANG JUANG: Fundamentals of Speech Recognition. Capítulo 8. PTR Prentice Hall. New Jersey, 1993.
- [2] L. Hernández Gómez, F. J. Caminero Gil, UNIVERSIDAD POLITÉCNICA DE MADRID, C. de la Torre Munilla, L. Villarrubia Grande, TELEFÓNICA INVESTIGACIÓN Y DESARROLLO, Estado del arte en Tecnología del Habla, Madrid, 2000
- [3] Erik McDermott, Member, Timothy J. Hazen, Member, “Discriminative Training for Large-Vocabulary Speech Recognition Using Minimum Classification Error”, IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 15, NO. 1, 2008
- [4] Nikos Chatzichrisafis, Vassilios Diakouloukas, Vassilios Digalakis, and Costas Harizakis, Gaussian Mixture Clustering and Language Adaptation for the Development of a New Language Speech Recognition System”, IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 15, NO. 3, 2008
- [5] Ascensión Gallardo-Antolín, Carmen Peláez-Moreno, *Member, IEEE*, and Fernando Díaz-de-María, *Member, IEEE* Recognizing GSM Digital Speech, IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 13, NO. 6, NOVEMBER 2005
- [6] Hong Kook Kim, *Senior Member, IEEE*, Richard V. Cox, *Fellow, IEEE*, and Richard C. Rose, *Senior Member, IEEE*, *Performance Improvement of a Bitstream-Based Front-End for Wireless Speech Recognition in Adverse Environments*, IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 10, NO. 8, NOVEMBER 2002
- [7] Ben Milner and Alastair James, *Member, IEEE*, Robust Speech Recognition Over Mobile and IP Networks in Burst-Like Packet Loss, IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 14, NO. 1, 2006
- [8] Michael L. Seltzer, *Member, IEEE*, Bhiksha Raj, *Member, IEEE*, and Richard M. Stern, *Member,*

IEEE, Likelihood-Maximizing Beamforming for Robust Hands-Free Speech Recognition, IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 12, NO. 5, SEPTEMBER 2004

[9] Heriberto Cuayáhuil Portilla, Técnicas para mejorar el reconocimiento de voz en presencia de habla fuera del vocabulario, Universidad de las Américas

[10] Miguel Angel Rodríguez Crespo, Ismael Cortázar Mágica, Daniel Tapias Merino, José Relaño Gil, Telefónica Investigación y Desarrollo, Estado del arte en tecnologías de voz, Marzo 2001

[11] Alina Almira, Gloria I. Alvarez, Antal A. Buss, Ernesto Castro, Diego L. Linares, Interfaz Oral para Lectura de Noticias en Internet, Octubre 2003

[12] Fernando Martinez, Gustavo Portale, Hernan Klein, Osvaldo Olmos, Reconocimiento de voz, Introducción a la Inteligencia Artificial

[13] Kimberlee A. Kemble, "An Introduction to Speech Recognition", Program Manager, Voice Systems Middleware Education IBM Corporation 2003

[14] Ben Milner, Alastair James, "Robust Speech Recognition Over Mobile and IP Networks in Burst-Like Packet Loss", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 14, NO. 1, 2008

[15] Javier Macías Guarasa, Proyecto de tesis doctoral, arquitecturas y metodos en sistemas de reconocimiento automático de habla de gran vocabulario, Grupo de Tecnología del Habla. Departamento de Ingeniería Electrónica Universidad Politécnica de Madrid.

[16] Speech Recognition Grammar Specification Version 1.0, <http://www.w3.org/2002/06/srgs-irp/> W3C 2002

[17] Jonathan Davidson, Brian Gracely and Jim Peters ,Voice over IP Fundamentals, Cisco Press

[18] Uyles Black, Internet Telephony: Call Processing Protocols, Prentice Hall

[19] Uyles Black , IP routing protocols: RIP, OSPF, BGP, PNNI & CISCO routing protocols, Prentice Hall Series in Advanced Communications Technologies

[20] Cysco Systems, Understanding Codecs: Complexity, Hardware Support, MOS, and Negotiation, Document ID: 14069

[21] William C. Hardy, Voip Service Quality Measuring And Evaluating Packet-Switched Voice, Mcgraw Hill

[22] Estándar IEEE 802.11, "IEEE Standard for Information technology-Telecommunications and information exchange between systems - Local and metropolitan area networks - Specific requirements: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications", 1999.

- [23] Jim Geier, Deploying Voice over Wireless LANs, Cisco Press
- [24] Michael L. Seltzer, Member, IEEE, Alex Acero, "Training Wideband Acoustic Models Using Mixed-Bandwidth Training Data for Speech Recognition", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 15, NO. 1, 2008
- [25] Francisco Javier Hernando Pericas, TECNICAS DE PROCESADO Y REPRESENTACION DE LA SEÑAL DE VOZ
- [26] Fundamentos del reconocimiento automático del habla
- [27] B. Resch. "Automatic Speech Recognition with HTK". Signal Processing and Speech Communication Laboratory.
- [28] Cisco Systems. Wireless quality-of-service deployment guide. Technical report, Cisco, 2003.
- [29] Szigeti T, Hattingh C. Quality of Service Design Overview.
- [30] Intel. Overcoming Barriers to High-Quality Voice over IP Deployments. 2007
- [31] J. Chou. "Design a successful VoWLAN system". Wireless Net DesignLine. Sep. 2005.
- [32] F. Mlinarsky. "Metrics and Methods Bring VoWLAN Success". Wireless Systems Design. Mar. 2005.

ANEXO A – CONFIGURACIÓN DE ASTERISK, D-ITG Y GENERACIÓN AUTOMÁTICA DE LLAMADAS

Evaluación de la Capacidad con ITG

Para realizar el proceso de evaluación por este camino, se instaló D-ITG, Distributed Internet Traffic Generator Versión 2.4 de la URL en cada uno de los equipos clientes en sus respectivos directorios y también en el equipo central

<http://www.grid.unina.it/software/ITG/download.php>.

Para las estaciones clientes se trabaja el programa `ITGSend` y para el equipo central `ITGRecv`. El ITG permite dos formas de calcular el retardo de un paquete, si es en una vía se requiere de un servidor NTP para sincronización de relojes y el receptor `ITGRecv` es el encargado de tomar los datos, si es el retardo de viaje redondo no se requiere el servidor NTP porque el transmisor `ITGSend` es el encargado de obtener los datos. Inicialmente se trató con NTP pero la sincronización se perdía rápidamente, además se consideró que la llamada sería en doble vía por lo que también era conveniente la segunda forma, por este motivo la toma de datos se tuvo en cada uno de las seis estaciones generadoras, cada una de las cuales generó un archivo (log) que al pasarlo al `ITGDec` se decodifica y presenta en formato *human-readable*. En este se encuentran los datos de retardo, *jitter* y pérdida de paquetes.

La condición límite se estableció basándose en los parámetros definidos en 4.1, de tal forma que la comunicación es inviable cuando se supere cualquiera de los límites de pérdidas, *jitter* o retardo mencionados.

En cada una de las estaciones clientes se construyó un script para que el ITGSend generara las llamadas que contenían las tramas de VoIP enviadas al equipo central, con los parámetros requeridos tales como cantidad y códec.

La sintaxis fue:

```
-a 192.168.1.2 -rp 10001 -t 60000 VoIP -x G.711.2 -h RTP
```

Como se trabajó con 6 PCs los puertos se escogieron colocando el primer número de acuerdo al cuarto byte de la dirección IP tal como aparece a continuación:

Cliente 1 - 192.168.1.101	-rp 10001-1000xx
Cliente 2 - 192.168.1.102	-rp 20001-2000xx
Cliente 3 - 192.168.1.103	-rp 30001-3000xx
Cliente 4 - 192.168.1.104	-rp 40001-4000xx
Cliente 5 - 192.168.1.105	-rp 50001-5000xx
Cliente 6 - 192.168.1.106	-rp 60001-6000xx

Se comenzó en forma secuencial con una conexión de VoIP con el Cliente 1, luego una segunda con el Cliente 2 y así sucesivamente hasta llegar al cliente 6, de esta manera seis llamadas activas significaba que los 6 clientes estaban generando llamadas. Como no se disponía de más clientes inalámbricos para que hubiera más llamadas se aumentó a que cada cliente generará dos o más llamadas, de tal forma que 12 llamadas implican que cada uno de los seis está generando dos y 18 que están generando tres.

El archivo de registro (log) que se obtuvo con cada uno se pasó al ITGDec para que lo decodificara y de esta forma conocer los datos de retardo, *jitter* y pérdida de paquetes. El límite se estableció por superar cualquier de ellos, pero como se mencionó anteriormente fue interesante ver que cuando se sobrepasaba la capacidad, todos estos límites eran considerablemente rebasados.

Se realizó primero con G.711, luego con G.729 y finalmente con G.723.

En el equipo central se ejecutó la aplicación lptraf para establecer ancho de banda

utilizado y paquetes por segundo tomando *screenshots* que comprobaran la utilización del canal y revisando si la entrada y/o salida era $n \times [BWc]$ donde n era la cantidad de conexiones y BWc el ancho de banda utilizado por cada códec. Para G.711 $BWc=64$, para G.729 $BWc=29$ y para G.723 $BWc=26.6$. Siempre se utilizó en tipo de protocolo RTP y no se utilizó VAD (Voice Activity Detection).

Antes de iniciar las pruebas se envió un archivo de 1.9 GB a cada cliente para ver si existía un promedio de velocidad relativo entre todos, con el fin de descartar problemas de desempeño en alguno de ellos, las velocidades relativamente fueron cercanas razón por la cual se consideró que todos se podían utilizar.

Evaluación de la Capacidad con Asterisk

Después de haber trabajado con el generador de tráfico, se decidió realizar el laboratorio con un servidor de VoIP generando llamadas reales entre los equipos y poder hacer una comparación y validación de resultados respecto al generador de tráfico D-ITG. Para llevar a cabo esta evaluación, fue necesario instalar la PBX IP Asterisk, realizar la configuración de los clientes y escribir algunos scripts que permitieran la generación automática de llamadas desde los clientes hacia el servidor, la respuesta de las llamadas en el servidor, y la reproducción automática de lado y lado de mensajes de voz pregrabados, de modo que se obtuvieran datos totalmente reales en una conversación de VoIP. Se tuvo el problema que si se usaban *softphones* como clientes, se tenía la limitación de tan solo una llamada por cliente incluso en los *softphones* que manejan llamadas simultáneas, pues estos realmente solo manejan una comunicación al tiempo y las demás llamadas las pone en espera, además de que cada llamada por *softphone* no se podría programar, ni ponerle a reproducir un mensaje automáticamente. Para solucionar este inconveniente, se instaló el servidor de VoIP asterisk en todos los clientes, ya que asterisk también puede operar como cliente, de modo que todo el trabajo, configuración y escritura de scripts se hizo para manejar las llamadas en asterisk. Adicionalmente al realizar la evaluación, se buscó tener una verificación “audible” al ser humano de la calidad de la voz, que demostrara cómo el aumento en el *jitter* y retardo, realmente afecta la percepción y la calidad de la comunicación. Para tal efecto, se instaló adicionalmente en uno de los clientes un *softphone*, con el cual se llamó al servidor cada vez que se quiso evaluar la calidad. Para más información sobre asterisk y su

configuración se puede visitar la URL www.asterisk.org.

A continuación se explican sólo las partes claves de la configuración, para facilitar la reproducción del experimento.

Configuración del Servidor

Sección de Clientes SIP “sip.conf”

En esta sección se declaró el abonado sip que se utilizó para monitorear la calidad de audio a través de parlantes

```
; Se declara la sección general
; Se declara la sección de abonado sip con un nombre
; cualquiera, para este caso [softphone1]
[softphone1]
type=friend                ;Permite hacer y recibir llamadas
context=ciclo              ;Podrá llamar a los números que se
                           ;incluyan en el contexto ciclo del
                           ;plan de marcado
host=dynamic               ;podra iniciar sesión desde cualquier
                           ;equipo
secret=telefono1          ;La contraseña
qualify=yes                ;monitorea la conexión y el retardo
dtmfmode=rfc2833         ;Detección de tonos estandar
relaxdtmf=yes              ;Facilita la detección de tonos
```

Sección de Plan de Marcado “extensions.conf”

En esta sección se declara la lógica y el flujo de la llamada

```
;Se declara la sección general
;Se declara el contexto que maneja las llamadas

[ciclo]                    ;Contexto Ciclo
exten=s,1,BackGround,demo-instruct
;Reproduce el mensaje demo-instruct que esta ya grabado      ;en el servidor
exten=s,2,Goto(ciclo,s,1)
;Crea un ciclo para que se repita el mensaje
;indefinidamente
```


exten=101,1,Goto(ciclo,s,1)
;cuando el softphone llama a 101 entra al ciclo

Sección de Clientes IAX "iax.conf"

;Se declara la sección general

;Se declara uno por uno los clientes a los que se ;conectará

[cliente1]

type=friend

;Puede hacer y recibir llamadas

host=192.168.1.100

;Se conectará desde esa IP

context=ciclo

;Podrá llamar a los número que se

;declaren en ese contexto

qualify=yes

;Se monitorea la conexión

;y el retardo

disallow=all

;No permite un códec diferente a

allow=ulaw

;g711u

[cliente2]

type=friend

;Puede hacer y recibir llamadas

host=192.168.1.101

;Se conectará desde esa IP

context=ciclo

;Podrá llamar a los número que se

;declaren en ese contexto

qualify=yes

;Se monitorea la conexión y el retardo

disallow=all

;No permite un códec diferente a

allow=ulaw

;g711u

;Se continúa con los demás clientes

Para cambiar de códec se puede modifica el allow por:

alaw Códec G.711a

ulaw Códec G.711u

g729 Códec G.729

g723 Códec G.723

gsm Códec GSM

ilbc Códec ilbc

Es de anotar que los códecs g729 y g723 no son libres y en ciertos ambientes, debe pagarse por ellos, asterisk los incluye solo en modo *passthrough*, de modo que para poder generar las llamadas y recibirlas, se debió compilar los códecs gratuitos para uso

académico disponibles en la página de intel www.intel.com e incluirlos en los módulos de asterisk.

Configuración del Cliente.

En los clientes solo fue necesario realizar una configuración en los clientes IAX, pues todas las llamadas se harían manejando este protocolo, solo se debe declarar el servidor al que irán conectados.

Sección IAX iax.conf

;Se declara la sección general

;Se declara el servidor que le llamaremos ap

[ap]

type=friend

host=192.168.1.2

context=ciclo

qualify=yes

disallow=all

allow=ulaw

trunk=no

;Para que cada llamada se haga como una

;llamada independiente. Si se coloca

;trunk=yes, se meten varias llamadas por una

;"conexión" que ahorra ancho de banda pero

;no es el caso real.

Una vez se haya iniciado asterisk en los clientes como en el servidor, y se haya hecho la configuración correcta, en la interfaz de línea de comandos CLI de asterisk, se pueden observar las conexiones de los clientes tal como aparece en la figura 4.

Generación de Llamadas

Una vez los clientes se encuentran conectados con el servidor, es necesario que se inicien las llamadas. Asterisk chequea constantemente un directorio en donde se pueden colocar scripts de llamadas para que el ejecute inmediatamente.

```

root@ryst18:~
Archivo  Editar  Ver  Terminal  Ira  Ayuda
[root@ryst18 root]# asterisk -rvvv
== Parsing '/etc/asterisk/asterisk.conf': Found
== Parsing '/etc/asterisk/extconfig.conf': Found
Asterisk 1.0.7, Copyright (C) 1999-2004 Digium.
Written by Mark Spencer <markster@digium.com>
=====
Connected to Asterisk 1.0.7 currently running on ryst18 (pid = 4255)
Verbosity is at least 3
ryst18*CLI> iax2 show peers
Name/Username      Host                Mask                Port                Status
cliente6           192.168.1.105      (S) 255.255.255.255  4569                OK (2 ms)
cliente5           192.168.1.104      (S) 255.255.255.255  4569                OK (3 ms)
cliente4           192.168.1.103      (S) 255.255.255.255  4569                OK (3 ms)
cliente3           192.168.1.102      (S) 255.255.255.255  4569                OK (3 ms)
cliente2           192.168.1.101      (S) 255.255.255.255  4569                OK (3 ms)
cliente1           192.168.1.100      (S) 255.255.255.255  4569                OK (3 ms)
demo/asterisk     216.207.245.47    (S) 255.255.255.255  4569                Unmonitored
ryst18*CLI>

```

Figura 4. Estado de conexión en el servidor

El script se detalla a continuación.

*;Script para generar llamadas debe ubicarse en /
;var/spool/asterisk/outgoing, tan pronto se copie el script ;en este directorio, se hará la ejecución.*

;Archivo sample.call

```

Channel: IAX2/ap                ; llama al servidor 'ap'
MaxRetries: 2                   ; hasta dos reintentos
RetryTime: 60                   ; Reintento cada 60 segundos
WaitTime: 30                    ; Esperara 30 segundos la respuesta
Context: ciclo                  ; La llamada llegará al conexto ciclo,
Extension: s                     ; extension s y
Priority: 1                       ; prioridad 1

```

Si se observa en las configuraciones anteriores, este script lleva la llamada a la ejecución del archivo pregrabado, que se ejecuta tanto en el servidor como en los clientes.

Para generar las llamadas se puede hacer un script que copie el archivo anterior las veces necesarias con diferente nombre cada vez, así:

```

#!/bin/bash
#mpstat -P 0 1 1 adiciona un retardo de un segundo entre #llamada y llamada generada además de mostrar el consumo
#de CPU
cp/tmp/sample.call /var/spool/asterisk/outgoing/sample1.call

```

```
mpstat -P 0 1 1
cp/tmp/sample.call /var/spool/asterisk/outgoing/sample2.call
mpstat -P 0 1 1
cp/tmp/sample.call /var/spool/asterisk/outgoing/sample3.call
mpstat -P 0 1 1
```

De esta forma se generan tres llamadas. En la figura 5, puede observarse el momento en el que se realizan varias llamadas desde un cliente

Es de anotar también que este script debe ejecutarse en cada cliente, pues es cada cliente quien genera la(s) llamada(s), en cuanto al servidor. también es importante aclarar, que para generar una llamada con diferentes características como cambio de códec, debe modificarse el archivo iax.conf tanto en los clientes como en el servidor cambiando allow=ulaw por otro códec.

Channel	Peer	Username	ID (Lo/Rem)	Seq (Tx/Rx)	Lag	Jitter	JitBuf	Format
IAX2/cliente6@client	192.168.1.104	cliente6	00001/00003	00003/00002	00000ms	0009ms	0000ms	ulaw
IAX2/cliente6@client	192.168.1.102	cliente6	00002/00003	00112/00111	00007ms	0007ms	0000ms	ulaw
IAX2/cliente6@client	192.168.1.100	cliente6	00004/00001	00204/00203	00000ms	0010ms	0000ms	ulaw
IAX2/cliente6@client	192.168.1.105	cliente6	00008/00002	00053/00052	00003ms	0004ms	0000ms	ulaw
IAX2/cliente6@client	192.168.1.101	cliente6	00010/00002	00137/00136	00003ms	0006ms	0000ms	ulaw
IAX2/cliente6@client	192.168.1.103	cliente6	00018/00002	00242/00241	00006ms	0006ms	0000ms	ulaw

0 active IAX channel(s)

Figura 5. Estado de cada llamada desde cada uno de los 6 clientes.

En la figura 6 se muestra el estado de 6 llamadas generadas hacia el servidor, una desde cada cliente

```
-- Executing Background("IAX2/cliente6@cliente6/22", "demo-instruct") in new stack
-- Playing 'demo-instruct' (Language 'en')
-- Accepting unauthenticated call from 192.168.1.105, requested format = 1, actual format = 1
-- Executing Background("IAX2/cliente6@cliente6/23", "demo-instruct") in new stack
-- Playing 'demo-instruct' (Language 'en')
-- Accepting unauthenticated call from 192.168.1.105, requested format = 1, actual format = 1
-- Executing Background("IAX2/cliente6@cliente6/24", "demo-instruct") in new stack
-- Playing 'demo-instruct' (Language 'en')
-- Accepting unauthenticated call from 192.168.1.105, requested format = 1, actual format = 1
-- Executing Background("IAX2/cliente6@cliente6/25", "demo-instruct") in new stack
-- Playing 'demo-instruct' (Language 'en')
-- Accepting unauthenticated call from 192.168.1.105, requested format = 1, actual format = 1
-- Executing Background("IAX2/cliente6@cliente6/26", "demo-instruct") in new stack
-- Playing 'demo-instruct' (Language 'en')
-- Accepting unauthenticated call from 192.168.1.105, requested format = 1, actual format = 1
-- Executing Background("IAX2/cliente6@cliente6/27", "demo-instruct") in new stack
-- Playing 'demo-instruct' (Language 'en')
-- Accepting unauthenticated call from 192.168.1.105, requested format = 1, actual format = 1
-- Executing Background("IAX2/cliente6@cliente6/28", "demo-instruct") in new stack
-- Playing 'demo-instruct' (Language 'en')
-- Accepting unauthenticated call from 192.168.1.105, requested format = 1, actual format = 1
-- Executing Background("IAX2/cliente6@cliente6/29", "demo-instruct") in new stack
-- Playing 'demo-instruct' (Language 'en')
-- Accepting unauthenticated call from 192.168.1.105, requested format = 1, actual format = 1
-- Executing Background("IAX2/cliente6@cliente6/30", "demo-instruct") in new stack
-- Playing 'demo-instruct' (Language 'en')
-- Accepting unauthenticated call from 192.168.1.105, requested format = 1, actual format = 1
-- Executing Background("IAX2/cliente6@cliente6/31", "demo-instruct") in new stack
-- Playing 'demo-instruct' (Language 'en')
-- Accepting unauthenticated call from 192.168.1.105, requested format = 1, actual format = 1
-- Executing Background("IAX2/cliente6@cliente6/32", "demo-instruct") in new stack
-- Playing 'demo-instruct' (Language 'en')
-- Accepting unauthenticated call from 192.168.1.105, requested format = 1, actual format = 1
-- Executing Background("IAX2/cliente6@cliente6/33", "demo-instruct") in new stack
-- Playing 'demo-instruct' (Language 'en')
```

Figura 6. Generación de llamadas desde un cliente

