



Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería en Ciencias y Sistemas

**“PORTALES DE VOZ: CAMINO HACIA INTERFACES HOMBRE-
MAQUINA MÁS NATURALES”**

SAMY EUNICE PINTO CASTAÑEDA

Asesorada por Inga. Ligia María Pimentel Castañeda

Guatemala, septiembre de 2004

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**“PORTALES DE VOZ: CAMINO HACIA INTERFACES HOMBRE-
MAQUINA MÁS NATURALES”**

TRABAJO DE GRADUACIÓN

PRESENTADO A JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

SAMY EUNICE PINTO CASTAÑEDA
ASESORADA POR: INGA. LIGIA MARÍA PIMENTEL CASTAÑEDA

AL CONFERÍRSELE EL TÍTULO DE
INGENIERA EN CIENCIAS Y SISTEMAS

GUATEMALA, SEPTIEMBRE DE 2004

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

DECANO	Ing. Sydney Alexander Samuels Milson
VOCAL I	Ing. Murphy Olympto Paiz Recinos
VOCAL II	Lic. Amahán Sánchez Álvarez
VOCAL III	Ing. Julio David Galicia Celada
VOCAL IV	Br. Kenneth Issur Estrada Ruiz
VOCAL V	Br. Elisa Yazminda Vides Leiva
SECRETARIO	Ing. Pedro Antonio Aguilar Polanco

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANO	Ing. Sydney Alexander Samuels Milson
EXAMINADOR	Inga. Ligia María Pimentel Castañeda
EXAMINADOR	Ing. Edgar Estuardo Santos Sutuj
EXAMINADOR	Ing. Luís Alberto Vettorazzi España
SECRETARIO	Ing. Pedro Antonio Aguilar Polanco

HONORABLE TRIBUNAL EXAMINADOR

Cumpliendo con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

“PORTALES DE VOZ: CAMINO HACIA INTERFACES HOMBRE-MAQUINA MÁS NATURALES”

Tema que me fuera asignado por la Escuela de Ingeniería en Ciencias y Sistemas de la Facultad de Ingeniería con fecha 20 de agosto de 2003.

Samy Eunice Pinto Castañeda

AGRADECIMIENTOS

A DIOS quien iluminó mi sendero para alcanzar mi ideal y obsequiárselo a Él quien lo merece todo.

A MIS PADRES Samuel Pinto y Alma Castañeda de Pinto; que mi triunfo sea para ellos una recompensa a sus múltiples esfuerzos.

A MI HERMANO Samuel Jonatán Pinto Castañeda; por todo su amor, paciencia y colaboración para alcanzar esta meta.

A MIS TÍOS David y Yolanda Hamm, por su apoyo constante en todas las áreas de mi vida.

A LA ASOCIACIÓN John Sarrin por su apoyo.

A LAS FAMILIAS Marroquín Morales y Samayoa Cárcamo; por brindarme siempre su apoyo.

A Ana María Nufio de Marroquín por su respaldo maternal.

A LA INGENIERA Ligia Maria Pimentel Castañeda; por su asesoría en la realización de este trabajo.

A La Universidad de San Carlos y a todos los catedráticos por haberme formado profesionalmente.

A todos mis compañeros y amigos quienes me brindaron su apoyo incondicional, especialmente a: Christian Chou-jo, Angela Villeda, Caridad Barrios, Edgar Veras, Oswaldo Cabrera, Vanessa Stwolinsky, Laura Lemus, Flavio Silvestre, Alan Ramírez, Eddy Velásquez, Daniel Quiñónez, Rubén Barrios, Arnulfo Hernández y René Alvarado.

ÍNDICE GENERAL

INDICE DE ILUSTRACIONES	V
GLOSARIO	VI
RESUMEN	XX
OBJETIVOS	XXIII
INTRODUCCIÓN	XXIV
1. EVOLUCIÓN DE LAS DIVERSAS TECNOLOGÍAS QUE COADYUVAN AL SURGIMIENTO DE LOS PORTALES DE VOZ	
1.1 Evolución de los medios para acceder a la red global.....	1
1.1.1 Páginas estáticas.....	2
1.1.2 Páginas dinámicas.....	3
1.1.3 Tecnología WAP.....	7
1.1.3.1 Ventajas de la tecnología WAP.....	10
1.1.3.2 Desventajas actuales del mercado WAP.....	10
1.1.4 Audiopáginas.....	11
1.2 Evolución de la Telefonía.....	12
1.2.1 Telefonía fija.....	12
1.2.2 Telefonía móvil.....	12
1.2.3 Tecnología de tercera generación (umts).....	12
1.2.4 Telefonía e Internet.....	13
1.3 Voz sobre IP	13
1.4 Tecnología de voz.....	18
1.4.1 Ingeniería lingüística.....	18
1.4.2 Accesibilidad y participación.....	21
1.4.3 El reconocedor de voz natural.....	22
1.4.3.1 Arquitectura del reconocedor de habla natural.....	23

1.4.3.2 Características que afectan al reconocedor de habla natural.....	24
1.4.4 El conversor texto a voz.....	25
1.4.4.1 Sistemas de conversión de texto a voz (tts).....	25
1.4.4.2 La conversión texto a voz avanzada.....	25
1.4.4.3 Aplicaciones.....	25
1.5 El ruido amenaza latente para los sistemas de reconocimiento de voz.	26
2 INTRODUCCIÓN A LOS PORTALES DE VOZ	
2.1 Definición de un portal de voz.....	27
2.2 Evolución histórica.....	29
2.3 Servicios de un portal de voz.....	31
2.3.1 Envío y recepción de e-mail.....	32
2.3.2 Agendas vocales.....	32
2.3.3 Directorio de empresas.....	33
2.3.4 Servicios de contenidos.....	33
2.4 Escenarios de operación para un portal de voz.....	34
2.5 Componentes de un portal de voz.....	36
2.6 Arquitectura de un portal de voz.....	37
2.6.1 Detalle de la plataforma de Acceso Vocal.....	38
2.7 Integración con la plataforma telefónica.....	45
2.8 Analogía con los portales visuales.....	45
2.9 Forma de uso.....	46
2.10 Objetivos de los portales de voz.....	48
2.11 Ventajas de los portales de voz.....	48
2.12 Desventajas de los portales de voz.....	50
2.13 Viabilidad.....	50
3 EL LENGUAJE ESTÁNDAR PARA EL ACCESO ORAL A INTERNET	
3.1 Historia.....	53
3.2 Definición.....	54

3.3	<i>World Wide Web Consortium</i>	55
3.4	Sintaxis.....	57
3.5	Ejemplos vistos desde la perspectiva de ejecución de un browser Vocal.....	65
3.6	Modelo arquitectónico del VXML y recursos necesarios para interactuar con el mismo.....	67
3.7	Contexto de ejecución de aplicaciones VoiceXML.....	69
3.8	Manejo de entradas y salidas de datos	70
3.9	Transcodificación de HTML o WML a VXML.....	75
3.10	WAP contra VoiceXML.....	78
3.11	Ventajas y desventajas de VoiceXML.....	79
4	FUNCIONAMIENTO DE LOS PORTALES INTERNACIONALES MÁS SOBRESALIENTES	
4.1	Portal de voz: TELLME.....	81
4.1.1	Reseña histórica.....	81
4.1.2	Descripción general de Tellme.....	81
4.1.3	Visión de Tellme.....	82
4.1.4	Logros de Tellme.....	82
4.1.5	Desarrollo de los servicios.....	82
4.1.6	Descripción de las soluciones más sobresalientes que Tellme ha creado para las empresas.....	83
4.2	Portal de voz: HEYANITA.....	84
4.2.1	Desarrollo de soluciones.....	85
4.2.2	Clientes y compañías con las que HeyAnita ha Trabajado.....	86
4.3	Portal de voz: YDILO.....	86
4.3.1	Reseña histórica.....	86
4.3.2	Descripción General de las soluciones proporcionadas por Ydilo.....	87
4.3.3	Objetivos de Ydilo.....	87
4.3.4	El trabajo con los clientes.....	89

4.3.5 Diseño y desarrollo de los servicios.....	89
4.3.6 Infraestructura de Ydilo.....	91
4.3.7 Diversidad de soluciones ofrecidas por Ydilo según el sector.....	92
4.3.8 Breve descripción de las soluciones más sobresalientes que Ydilo a desarrollado para empresas y eventos.....	95
5. EXPECTATIVAS PARA EL FUTURO	
5.1 Fuerzas que demandan la Web de voz.....	99
5.2 Generalidades y expectativas del futuro de los portales de voz.....	99
5.2.1 Generalidades.....	99
5.2.2 Expectativas de los portales de voz.....	100
5.2.3 Perfil de los desarrolladores de los portales de voz....	101
5.3 El futuro del desarrollo tecnológico.....	101
5.3.1 Desarrollo tecnológico.....	101
5.3.2 Protagonistas en la Web de voz.....	102
5.4 Los portales de voz y el comercio electrónico.....	104
5.4.1 V-Commerce.....	104
5.4.2 Seguridad.....	105
5.5 Desarrollo y expectativas de portales de voz en Guatemala.	105
5.5.1 Viabilidad técnica, económica y social.....	105
5.6 Primer portal de voz en Guatemala.....	108
CONCLUSIONES	115
RECOMENDACIONES	118
REFERENCIAS	121
BIBLIOGRAFIA	122
APENDICE	125

ÍNDICE DE ILUSTRACIONES

FIGURAS

1.	Una red WAP	8
2.	Redes de telefonía actual vrs Red sobre IP	14
3.	Elementos de una red VoIP	17
4.	Diagrama a bloques de un sistema de dialogo	23
5.	Evolución histórica para el acceso a Internet	30
6.	Escenarios de operación	35
7.	Reconocedor de habla	33
8.	Plataforma de acceso vocal a Internet	39
9.	Diseño del HW y SW del servidor de aplicaciones	41
10.	Arquitectura básica de las aplicaciones escritas en VoiceXML	68
11.	Contexto de ejecución de aplicaciones VoiceXML	70
12.	Transcodificación dinámica de contenidos WEB y WAP	77
13.	Protagonistas en la Web de voz	102

TABLAS

I.	Internet tradicional contra Internet por voz	46
II.	Estructura Básica de VoiceXML	58
III.	Aplicación VoiceXML con transito entre forms	63
IV.	WAP contra Aplicación en VoiceXML	78

GLOSARIO

Applets	Aplicación escrita y compilada en Java, su característica principal es que se ejecutan en el equipo cliente.
ASP	<i>Active Server Pages.</i> Son paginas que se ejecutan en el servidor y en las cuales se utiliza algún lenguaje de <i>scripting</i> , como VBScript o JavaScript.
ASR	<i>Automatic Speech Recognition.</i> Reconocimiento de voz automático, tecnología capaz de reconocer órdenes y nombres directamente de la voz humana.
AT&T	Compañía Estadounidense de telecomunicaciones. Una división de esta compañía, la <i>Bells Lab</i> .
ATM	<i>Asynchronous Transmission Mode.</i> Modo de transmisión asíncrona. Sistema de transmisión de datos usado en banda ancha para aprovechar al máximo la capacidad de una línea.
BD	Término utilizado para referirse a una base de datos también se utiliza BDD.
Browser Vocal	En español navegador de voz. Es el elemento utilizado de puente entre los contenidos Web y el usuario.

Call Center	Esta palabra se traduce al español como Centro de Llamadas que es un sistema integrado de telefonía y computación orientado a potenciar las 3 labores más importantes de una empresa, por medio de una comunicación telefónica las cuales son, la adquisición de clientes, el mantenimiento de clientes y el cobro a través del sistema telefónico.
CDMA	<i>Coda Division Multiple Access</i> Acceso múltiple mediante división de código. Sistema utilizado en las comunicaciones móviles de acuerdo con el estándar US (IS 95) en el intervalo de frecuencias entre los 800 y los 1900 Mhz.
CDPD	Cellular Digital Packet Data, paquetes de datos digitales celulares. Servicio de datos celulares.
CGI	<i>Common Gateway Interface</i> . Interfase de acceso común. Programas usados para hacer llamadas a rutinas o controlar otros programas o bases de datos desde una página Web. También pueden generar directamente HTML.
CHAT	Charla. Servicio que permite a cualquier persona comunicarse con sus semejante mediante mensajes escritos en tiempo real.
Conmutador	Dispositivo externo que se utiliza para activar un dispositivo entre varios conectados al mismo: módem, impresora, etc.

CSS	<i>Cascading Style Sheets</i> . Hojas de estilo en cascada, utilizadas para incluir en ellas tipos de letras, márgenes, etc.
DECT	<i>Digital Enhanced Cordless Telecommunications</i> . Norma de armonización de las comunicaciones domésticas (o profesionales) a través de terminales sin cables. Los aparatos DECT utilizan comunicación digital sin cables para que puedan ofrecer una vasta gama de funciones y una calidad sonora muy buena.
DHTML	<i>Dynamic HTML</i> . Utilizado para generar contenidos dinámicamente y responder de esta forma a acciones del usuario.
DNS	<i>Domain Name System</i> . Sistema de nombres de dominio. Base de datos distribuida que gestiona la conversión de direcciones de Internet expresadas en lenguaje natural a una dirección numérica IP.
DTD	<i>Document Type Definition</i> . Documento de definición de tipo que permite ver si cierto archivo concuerda con la estructura correspondiente.
DTMF	Dual Tone Multifrequency. Multifrecuencia de doble tono. Son los tonos que se utilizan en telefonía para marcar un número telefónico.
E-Commerce	Es la realización de negocios en Internet, no solo la compra y venta, sino también proporcionar servicio a clientes y colaborar con asociados de negocios.

EPOC	Sistema operativo desarrollado por Psion, utilizado principalmente para dispositivos móviles. Actualmente evolucionó con el nombre de Symbian.
ETHERNET	Estándar IEEE 802.3 para redes de contención. Utiliza una topología en bus o estrella, y depende de la forma de acceso conocida como Acceso múltiple con detección de portadora y colisiones para regular el tráfico en la línea de comunicaciones. Los nodos de la red se vinculan mediante cable coaxial, cable de fibra óptica o cable de par trenzado. Los datos se transmiten en tramas de longitud variable que contienen información de entrega y de control.
Frame Relay	Protocolo de enlace mediante circuito virtual permanente, muy usado para dar conexión directa a Internet.
FTP	<i>File Transfer Protocol</i> . Protocolo de transferencia de archivos, utilizado para transferir archivos a través de una amplia variedad de sistemas. Uno de los protocolos más usados en Internet.
Gatekeeper	Centro de toda organización de voz sobre IP, que sustituye a la actual central telefónica.
Gateway	Pasarela que permite la conversión de unos protocolos en otros, consiguiendo así la conexión de redes que de otro modo no sería posible.

GPRS	<i>General Packet Radio Services</i> Servicio de Radio de Paquetes Generales; tecnología vinculada a paquetes que posibilita una Internet móvil de alta velocidad y otras comunicaciones de datos.
GSM	<i>Global System for Mobile Communications</i> Sistema Global de Comunicaciones Móviles. Sistema digital de telecomunicaciones principalmente usado para telefonía móvil. Existe compatibilidad entre redes por tanto un teléfono GSM puede funcionar teóricamente en todo el mundo.
Handheld	Cualquier computadora o dispositivo que quepa en la mano del usuario, incluyendo palms, teléfonos portátiles, pagers, mini laptops, etc.
Hipermedia	Extensión del concepto de hipertexto para la inclusión de multimedia (sonido, gráficas y vídeo).
Hipertexto	Describe un tipo de funcionalidad de exploración en línea interactiva. Los vínculos (direcciones URL) incrustados en palabras o frases permiten al usuario seleccionar texto y mostrar inmediatamente información relacionada y material multimedia.
HTML	<i>Hypertext Markup Language</i> . Lenguaje de "etiquetas" en el que se asigna formato a las páginas Web y se distribuye la información.

HTTP	<i>HyperText Transfer Protocol</i> . Protocolo de transferencia de hipertexto. Método mediante el que se transfieren documentos desde el sistema host o servidor a los exploradores y usuarios individuales.
HUB	Dispositivo de red que concentra todas las conexiones de los dispositivos de una red. A cada ranura se le llama puerto (Slot), y algunos tienen ranuras especiales para conectarlos usando crossovers en serie llamadas Uplink.
iDEN	Siglas de <i>Integrated Digital Enhanced Networks</i> . Esta solución inalámbrica de Motorola está diseñada para los usuarios de negocios móviles que necesitan un acceso inmediato a la información y demás usuarios, sin tener que llevar diversas piezas de equipamiento. Combina las capacidades de un teléfono celular, radio de dos direcciones, pager alfanumérico y módem data/fax en un solo dispositivo portátil.
IP	<i>Internet Protocol</i> Protocolo de Internet. Dirección de 32 bits del protocolo Internet asignada a un host. La dirección IP tiene un componente del host y un componente de la red.
IRC	<i>Internet Relay Chat</i> . Canal de Chat de Internet. Sistema para transmisión de texto multiusuario a través de un servidor IRC. Usado normalmente para conversar on-line.

ISP	<i>Internet Service Provider</i> . Proveedor de Servicios Internet.
IVR	IVR o VRU <i>Interactive Voice Response</i> Aplicación de software que acepta una combinación de entradas de voz y tonos telefónicos, para proveer servicios automatizados de respuesta adecuados tales como correo de voz, fax, correo electrónico, etc. Incluyen la capacidad de enlazarse con bases de datos para entregar contenido tales como datos de encuestas, transacciones bancarias, etc. Incluso, pueden grabar la voz del la otra persona para posterior procesamiento.
JavaOS	Sistema Operativo de Sun basado en la tecnología Java y dispuesto para integrarse perfectamente con él.
JavaScript	Programa escrito en el lenguaje script de Java que es interpretado por la aplicación cliente, normalmente un navegador (<u>Browser</u>).
JSP	<i>Java Server Pages</i> Utilizado para crear páginas que se ejecutan en el servidor y regularmente se utiliza JavaScript.
LAN	(Local Area Network) Red de Area Local. Conjunto de ordenadores conectados entre sí y geográficamente próximos (Mismo edificio, habitación, etc.)

LDAP	LDAP (<i>Lightweight Directory Access Protocol</i>) es el protocolo utilizado para la comunicación entre un programa que efectúa una llamada (que se ejecuta en un nodo como, por ejemplo, un servidor de comercio) y un nodo de directorio. En el nodo de directorio basado en LDAP se guarda información acerca de temas como los servicios y las personas.
MMS	El servicio de mensajes multimedia es un nuevo estándar que se está desarrollando para su uso en terminales móviles avanzados. El servicio permite la transmisión en tiempo no real de diversos tipos de contenido multimedia como imágenes, videoclips, etc.
Pager	Pequeño dispositivo de telecomunicación donde se reciben mensajes que aparecen escritos en un display. La comunicación se establece por teléfono y también por e-mail y es de una sola vía: el usuario debe responder el llamado comunicándose por otro medio.
PalmOS	Sistema operativo utilizado en los dispositivos Palm™ y en los de otros fabricantes, como Sony o Handspring.
PBX	Comúnmente llamado conmutador, es el sistema de intercambio de líneas telefónicas.
PDAs	Personal Digital Assistant. Asistente personal digital, nombre ordinario que se le da a las agendas electrónicas de bolsillo. Y que se encarga de atender a un usuario concreto en tareas como búsquedas de información o selecciones.

PERL	Lenguaje para manipular textos, ficheros y procesos. Con estructura de script. Desarrollado por Larry Wall, es multiplataforma.
PHP	Preprocessed Hypertext Pages. Lenguaje de programación destinado a la Red, y que una vez interpretado por el servidor Web genera código HTML.
PIN	Personal Identification Number. Número personal de Identificación. Número secreto asociado a una persona o usuario de un servicio mediante el cual se accede al mismo. Se podría decir que es una contraseña o clave.
POP3	Conocido protocolo que se utiliza para recibir mensajes de correo electrónico. Suelen utilizarlo los proveedores de servicios Internet. Los servidores POP3 permiten tener acceso a una sola bandeja de entrada, a diferencia de los servidores IMAP, que proporcionan acceso a múltiples carpetas en los servidores.
PROXY	El Proxy es un servidor que conectado normalmente al servidor de acceso a la WWW de un proveedor de acceso va almacenando toda la información que los usuarios reciben de la WEB. Por tanto, si otro usuario accede a través del Proxy a un sitio previamente visitado, recibirá la información del servidor Proxy en lugar del servidor real.

RDSI	Red Digital de Servicios Integrados. (en inglés ISDN) Servicio telefónico digital de alta velocidad que puede aumentar sustancialmente la velocidad de conexión a Internet o a una red de área local (LAN) corporativa. El servicio ISDN (RDSI) puede funcionar a 128 kilobytes por segundo (Kbps).
Red	Sistema de elementos interrelacionados que se conectan mediante un vínculo dedicado para proporcionar una comunicación local o remota (de voz, vídeo, datos, etc.) y facilitar el intercambio de información entre usuarios con intereses comunes.
Router	Encaminador. Dispositivo que une entre sí dos redes, de forma que la información que no va dirigida a la otra red, no pasa a ella. Dispositivo conectado a dos o más redes que se encarga únicamente de tareas de comunicaciones.
RSVP	Protocolo que se encarga de dar prioridad a los paquetes de voz cuando hay congestión en un router.
RTB	Red Telefónica Básica: Es una línea telefónica analógica de cobre con un solo canal de comunicación.
RTP	RTP/RTCP Real Time Protocol/ Real Time Control Protocol. Protocolo de Tiempo Real. Protocolo utilizado para la transmisión de información en tiempo real como por ejemplo audio y video en una videoconferencia.

Servlets	Módulos que permiten sustituir o utilizar el lenguaje Java en lugar de programas CGI.
Smartphone	La próxima generación de teléfonos móviles que incorporará las funcionalidades de un dispositivo Palm.
SMS	<i>Short Message Services</i> . Servicio disponible en redes digitales que permite enviar y recibir hasta 160 caracteres en un teléfono móvil, a través del centro de mensajes del operador de la red.
SONET	Red óptica de datos de muy altas velocidades con sincronización en modo full duplex.
TDMA	<i>Time Division Multiple Access</i> . Tecnología que permite llevar servicios inalámbricos digitales utilizando y manteniendo diversos flujos de información de manera independiente a través de un mismo canal de comunicación.
TETRA	<i>Terrestrial Trunked Radio Access</i> , estándar original europeo para sistemas de radio móvil terrestre (LMR).
The Kelsey Group	<i>The Kelsey Group</i> , una firma de investigación de mercado de Princeton (Nueva Jersey) Estados Unidos.
Transcodificación	Término con el que se hace referencia al proceso para lograr que una aplicación de comercio electrónico existente pueda visualizarse en las distintas modalidades de pantalla de los dispositivos móviles.

La transcodificación consiste en convertir datos y aplicaciones escritas en lenguajes Web estándar, como HTML y XML, en formatos tales como Wireless Markup Language, XHTML o VXML.

TTS	<i>Text to Speech</i> . Sistema de conversión de texto a voz también llamado STT.
UMTS	Universal Mobile Telecommunications System: sistema de telecomunicaciones móviles universal. <i>Es el protocolo de la tercera generación de teléfonos celulares.</i>
URI	<i>Universal/Uniform Resource Identifier</i> . Identificador Uniforme/Universal de Recurso.
URL	Uniform Resource Locator) Formato de las direcciones de sitios que muestra el nombre del servidor en el que se almacenan los archivos del sitio, la ruta de acceso al directorio del archivo y su nombre.
Voice Browser	Término utilizado para referirse al navegador de voz
VoiceXML	Un nuevo estándar que permite el acceso al contenido Web a través del teléfono. VoiceXML utiliza XML para representar el flujo de la llamada y del diálogo, y permite tanto el acceso, la navegación y la recuperación de contenidos de sitios Web que cumplan este estándar utilizando cualquier teléfono, incluyendo los móviles.

VoIP	Término que se utiliza para abreviar “Voz sobre IP” tecnología que permite la transmisión de la voz a través de redes IP, Internet normalmente. La telefonía IP es una aplicación inmediata de esta tecnología.
Voz Sintetizada	Es el resultado del conversor de texto a voz, mediante el cual el sistema se comunica con el usuario a través de voz.
VXML	Abreviatura de VoiceXML.
W3C	Word Wide Web Consortium. Grupo que consolida las especificaciones del lenguaje VoiceXML.
WAP	<i>Wireless Application Protocol</i> Protocolo Inalámbrico de Aplicación, tecnología que permite enviar y recibir información de Internet desde un teléfono móvil.
WAV	Formato Windows, y también la extensión de los ficheros de audio.
WEB	Sistema para explorar Internet mediante hipervínculos. Cuando se utiliza un explorador Web, el Web aparece como una colección de texto, imágenes, sonidos y películas digitales.
Windows CE	Sistema operativo de Microsoft para handhelds y otros dispositivos móviles.

WML	<i>Wireles Markup Languaje</i> . Es el lenguaje de marcas inalámbrico.
WWW	World Wide Web Red mundial amplia, conocido también como: W3 ó el Web.
XML	Siglas de "Extensible Markup Language". Una herramienta de formato de página Web que es utilizada para crear páginas Web y para desarrollar aplicaciones para Palms.
XSL	Hoja de estilo que controla cómo se presentan los datos de un archivo XML

RESUMEN

Internet se ha convertido en un medio de comunicación muy utilizado en la actualidad. Diversidad de tecnologías se han desarrollado para permitir que esta red funcione de una mejor manera. Hoy en día se puede dar un vistazo al pasado y observar varias herramientas, como las páginas estáticas y las dinámicas, las cuales en su momento fueron las únicas que hicieron posible a las personas intercambiar todo tipo de información a través de la red. Paralelo a este progreso se ha visto la evolución que la telefonía ha tenido y que ha permitido que las comunicaciones no sean manejadas solamente por líneas fijas sino también por la tecnología móvil. Todos estos avances provocaron la necesidad de conseguir medios para un acceso más cómodo y portátil a Internet conduciendo al desarrollo de la tecnología WAP. Sin embargo, la evolución continua dando paso al surgimiento de tecnologías como UMTS y Voz sobre IP, las cuales, a su vez, permiten que las comunicaciones integren más y mejores servicios.

La tecnología de voz también ha mostrado grandes avances, entre los que destacan el reconocedor de voz y el conversor texto a voz. Todos estos avances han hecho posible que el ideal de acceder a la Web a través de la voz y sin necesidad de grandes máquinas o teléfonos sofisticados, sino solo con un teléfono convencional, sea hoy posible, pues es con la unión de todos estos avances es decir tanto en la telefonía como en las herramientas para Web, que surgen los portales de voz.

Los portales de voz permiten acceder a la red global mediante la manera más fácil y natural ya que lo único que los usuarios necesitan es la voz y el uso de un teléfono común. Esto da lugar a una navegación vocal que pretende relacionar el mundo Internet con el mundo de la voz. Sin

embargo, como todo desarrollo tecnológico este resulta de mayor utilidad para brindar servicios como: noticias, localización geográfica de diversos tipos de lugares, mensajería unificada, agendas vocales, envío y recepción de mail, entretenimiento, servicios de directorio. Aunque también pueden ser incorporados otros.

VoiceXML es el lenguaje estándar que se ha creado para la construcción de las audiopáginas, las cuales dan vida a los portales de voz. Este es un lenguaje extensible de marca de voz. Sigue todas las reglas sintácticas de su antecesor, XML. El grupo que se encarga de normalizar todos los aspectos de este lenguaje y permitir que el mismo se consolide lleva por nombre W3C (World Wide Web Consortium). VoiceXML es similar a HTML y a WML lo cual permite que las mismas personas que han trabajado construyendo sitios con HTML o WML puedan ahora crear las audiopáginas.

El funcionamiento de esta nueva forma de acceder a Internet ya puede ser observado en varios países. Entre los portales más sobresalientes en la actualidad se pueden mencionar: Tellme (primer portal de voz), Heyanita e Ydilo (los primeros en USA y el último en España). Estas tres empresas han demostrado lo fácil, rápido y cómodo que es ahora acceder a contenido Web mediante la voz. Estos portales de voz se han enfocado en ofrecer soluciones a diversidad de empresas, dándoles a conocer cómo pueden ofrecer sus productos o servicios mediante diálogos hombre-máquina, lo cual puede otorgarles mayores beneficios económicos.

En Guatemala, existe una empresa que cuenta con un portal de voz, claro que por el momento se dedica a prestar los servicios básicos, como los descritos arriba, y pueden ser utilizados solamente por sus clientes. Sin embargo, la persona que así lo desee ya puede experimentar en este país lo que es el acceso a contenido Web mediante la voz.

El futuro de los portales de voz es bastante prometedor, principalmente porque cuenta con una interacción basada en la voz. En la actualidad se considera que este nuevo medio de acceso a la Web será en poco tiempo una buena opción para el comercio. No obstante, es necesario confirmar aspectos relacionados con la seguridad en las transacciones comerciales, las cuales, en este caso, serán orales. Con todo ello se espera que en el futuro el interés por contar con una audiopágina en uno de los portales de voz sea similar al que hoy existe por tener una página Web.

OBJETIVOS

General

Dar a conocer en forma amplia un nuevo canal de interacción basado en comunicar mediante el teléfono común, celular o de cabina telefónica los diversos contenidos online, utilizando como base de estos portales el lenguaje de programación VoiceXML

Específicos

1. Conocer la evolución histórica del acceso a Internet
2. Explicar qué son los portales de voz, como funcionan, cuales son sus componentes y las diversas ventajas que ofrecen
3. Investigar el lenguaje VoiceXML como estándar sobre el que están soportados los portales de voz
4. Presentar algunos ejemplos de portales de voz en otros países, y el impacto que los mismos han tenido
5. Mostrar las expectativas desarrollo y uso de estos portales en Guatemala

INTRODUCCIÓN

El deseo del hombre por interactuar con las máquinas de forma similar a como lo hace con su especie, ha dado lugar a numerosas investigaciones y descubrimientos de tecnologías que han contribuido al desarrollo de la humanidad.

Desde la aparición de la computadora, el avance tecnológico se ha incrementado considerablemente. Día tras día se conocen nuevas herramientas que, junto a la PC, proveen mejores soluciones al entorno en el que interactuamos.

El ser humano es un ente sociable, por lo que sus prioridades siempre serán buscar mejores y sofisticadas formas de comunicación.

Internet se ha convertido en la fuente universal de información y comunicación, ya que en ella es posible encontrar contenidos de diversa índole, así como participar en varios tipos de comunicación con otras personas (chats, foros, videoconferencias etc.).

Los medios para presentar contenido Web, como para acceder a la red global, han evolucionado grandemente. Primero fueron las páginas estáticas, luego las dinámicas. Después surgió WAP como una alternativa para acceder a Internet mediante dispositivos móviles que apoyen esta tecnología.

Las investigaciones sobre incorporar una nueva opción para transmitir la voz, han dado surgimiento a la voz sobre IP, la cual pretende transmitir la voz mediante las redes de datos. Esto reemplazará la plataforma telefónica

que actualmente utilizamos dando paso a una moderna comunicación que abrirá las puertas a muchos otros proyectos.

La posibilidad de interactuar con los sistemas informáticos mediante la voz, sin tener que recurrir al teclado, constituye el principal objetivo en la actualidad, ya que esto permitirá una interacción más natural y flexible con los sistemas informáticos.

XML ha proveído los elementos necesarios para dar origen a VoiceXML lenguaje que se está comenzando a utilizar como estándar para el desarrollo de audiopáginas.

Toda esta evolución de tecnologías ha permitido el nacimiento de “Los portales de voz”, los cuales se encargan de presentar los contenidos Web mediante audiopáginas.

Las audiopáginas permitirán que en el futuro la voz sea la que dirija nuestra relación con la Web y lo podremos hacer utilizando cualquier teléfono convencional o móvil.

En su día la PC lo cambió todo y ahora le toca el turno a una nueva revolución: la voz.

1. EVOLUCIÓN DE LAS DIVERSAS TECNOLOGÍAS QUE COADYUVAN AL SURGIMIENTO DE LOS PORTALES DE VOZ

1.3 Evolución de los medios para acceder a la red global

Las comunicaciones vía Internet son hoy en día una necesidad indiscutible para un buen número de personas. Por ello las diversas instituciones, compañías, organizaciones, independientemente del sector que ocupen, saben que tienen que tener presencia en la Web y hacer frente a un creciente número de clientes que demandan seguridad, fiabilidad y novedades que mantengan su interés.

Los servicios básicos de toda aplicación en línea, como el correo electrónico, IRC o Chat, FTP, y el propio Web, son el punto de partida; sin embargo, en la actualidad ya se puede recibir una gran cantidad de servicios dinámicos e interactivos presentes generalmente en los grandes portales.

En sus inicios al sistema en el que está construido la Web se le llamaba hipertexto, donde lo que predominaba eran páginas conectadas con enlaces. Esto permitía a los usuarios navegar fácilmente, pues con solo ir pulsando enlaces se lograba acceder a la información de interés. Con el avance de todo este sistema, la Web ya no solo presenta textos y enlaces, sino también imágenes, videos, sonido y cualquier otro tipo de presentación, por lo que ahora se utiliza un nuevo término, que es hipermedia, dando a conocer que la Web permite contenidos multimedia.

1.3.1 Páginas estáticas

Este tipo de páginas fueron las que en un inicio poblaron los servidores Web, creados en su mayoría en el conocido lenguaje HTML. Este es un formato básico para presentar el contenido estático. Indica dónde colocar cada texto, imagen o video y la forma que estos tendrán al estar colocados en la página.

Este formato fue creado con objetivos puramente divulgativos, sin pensar que en un futuro existirían diversos usos. En síntesis, el HTML es una serie de etiquetas que se utilizan para definir la forma o estilo que se quiere aplicar al documento.

Las hojas de estilo en cascada (conocidas por sus siglas en inglés CSS, *Cascading Style Sheets*), también se comenzaron a utilizar. Con las mismas es posible modificar de manera rápida y coherente el aspecto de las páginas. Son consideradas un estándar amplio, con unas especificaciones y posibilidades muy grandes. CSS permite incluir cosas que no se podían utilizando solamente HTML, como márgenes, tipos de letra, fondos, colores, etc. Estas hojas de estilo en cascada se escriben dentro del código HTML de la página Web; sin embargo, para casos avanzados es posible escribir en un archivo aparte y luego enlazar la página con ese archivo.

XML (*Xtensible Markup Language*): Es un lenguaje de marcas (basado en texto puro), vagamente parecido al HTML, pero que solo contiene información, contenidos puros (no información sobre presentación). El contenido de un fichero XML se representa utilizando otros recursos (como hojas XSL). Además, los elementos de un fichero XML pueden definirse a voluntad por el usuario; se puede utilizar XML para representar información muy diversa.

La estructura válida de un fichero XML puede ser descrita mediante un DTD (*Document Type Definition*) que permite verificar si cierto fichero se ajusta a la estructura correspondiente. Existe otra técnica, *XML Schemas*, con fines similares. Además, para presentar información, el XML se utiliza como formato de intercambio de la información entre diferentes plataformas o aplicaciones.

XSL es básicamente, una especie de hojas de estilo que permiten controlar cómo se presentan los datos de un archivo XML (ya que XML no lleva información alguna sobre presentación). Es frecuente que un fichero XML se combine con un XSL para dar lugar al HTML que se ve en el navegador.

1.3.2 Páginas dinámicas

Las páginas dinámicas son las que en la actualidad permiten intercambiar información con el usuario de forma transparente, o simplemente presentar cualquier efecto especial u otra funcionalidad para lo cual es necesario utilizar otros lenguajes de programación, aparte del ya mencionado HTML. Por lo general, al escribir una página dinámica el código de los otros lenguajes de programación se incluye embebido dentro del mismo código HTML (ya que este se sigue considerando como la base para el desarrollo Web).

Existen varias razones para construir páginas dinámicas, entre las que se encuentran: la simple vistosidad que se puede alcanzar, debido a que se pueden hacer presentaciones más entretenidas de las que se consiguen utilizando solo HTML; la posibilidad de automatizar un proceso y, con ello, permitir que la página se encargue de las actualizaciones necesarias.

Mediante este tipo de páginas el usuario puede enviar sus datos al servidor Web, el cual las procesa y devuelve el resultado solicitado. Para poder llevar a cabo esta interacción se manejan básicamente 2 conceptos:

Cliente: llamado así el ordenador donde el usuario se conecta a Internet.

Servidor: ordenador donde se encuentran alojados los diversos contenidos Web a los que se puede acceder.

Las páginas dinámicas cliente son las que se procesan en el cliente, por lo que todo el procesamiento de la funcionalidad debe estar soportada por el navegador. Por lo general los usos que se le dan son efectos especiales para Web, presentaciones en las que se pueden mover objetos por la página, control de formularios, cálculos, etc. El código que se necesita para crear estos efectos y funcionalidades también se incluye dentro del mismo archivo HTML y es llamado SCRIPT. Los lenguajes de programación en los que generalmente se escriben este tipo de páginas son Javascript y Visual Basic Script.

Con las páginas del servidor se puede hacer todo tipo de aplicaciones Web, como agendas, foros, estadísticas, chats, juegos, sistemas de documentación, etc. Se utiliza principalmente cuando se tiene que acceder a información centralizada, ubicada en una base de datos en el servidor, y que por razones de seguridad los cálculos no se pueden realizar en el ordenador del usuario.

Para la mayoría de aplicaciones Web se debe tener acceso a muchos recursos externos al cliente, principalmente BDD, por lo cual es importante utilizar páginas dinámicas de servidor. Al igual que las páginas del cliente, se escriben en el mismo archivo HTML. Cuando una página es solicitada por parte de un cliente, el servidor ejecuta los scripts y se genera una página resultado, que solamente contiene código HTML. Este resultado final es el que se envía al cliente y puede ser interpretado sin lugar a errores ni incompatibilidades, puesto que solo contiene HTML.

Para escribir páginas dinámicas de servidor existen varios lenguajes, como: CGI's, comúnmente escritos en Perl, Páginas ASP, PHP, y JSP.

Las ventajas de este tipo de programación son que el cliente no puede ver los scripts, ya que se ejecutan y transforman en HTML antes de enviarlos, y que son independientes del navegador del usuario, ya que el código que reciben es HTML, fácilmente interpretable.

DHTML (*Dynamic HTML*) es utilizado para generar contenidos dinámicamente, responder a acciones del usuario (clics y demás), controlar el navegador (menús, sub-ventanas), controlar elementos CSS sin necesidad de recargar la página entera. Básicamente, se usan diversas capas, que se pueden ocultar, mostrar y mover.

Applets es un elemento del lenguaje Java que permite escribir pequeños programas que gobiernan el comportamiento de una región de una página HTML. Gracias a esto se pueden hacer páginas interactivas.

La característica principal de los Applets es que se ejecutan en el equipo cliente, dentro del navegador. Por ello se necesita la máquina virtual de Java instalada en dicho equipo. Además, tiene severas restricciones de seguridad. Por ejemplo, no puede acceder al disco duro del equipo cliente.

VBScript es un lenguaje de programación con sintaxis similar al Visual Basic, utilizado en las páginas HTML como lenguaje de scripting en el lado cliente (interpretado por el navegador). Sin embargo, tiene una desventaja y es que solo funciona en navegadores de Microsoft (Internet Explorer).

JavaScript es un lenguaje de programación utilizado en las páginas HTML como lenguaje de scripting (interpretado por el navegador). Permite que las páginas sean interactivas, o incluso que se modifiquen dinámicamente en el lado cliente. Contrariamente a lo que pueda parecer, no

tiene nada que ver con Java, y su sintaxis es solo vagamente similar a este (también es vagamente similar a la del lenguaje C). El JavaScript también se puede usar en el lado servidor, aunque su uso en el cliente es lo más habitual.

CGI es el sistema de mayor antigüedad que existe para las páginas dinámicas de servidor. Se trata de simples programas (que pueden ser ejecutables binarios, programas Perl, etc.). Un programa CGI recibe ciertos parámetros por dos posibles vías: una, mediante variables de entorno (método GET) y otra mediante la entrada estándar (método POST). A su vez, el programa realizará el proceso que deba y escribirá en su salida estándar la página HTML resultante. Un CGI se pueda implementar prácticamente en cualquier lenguaje y sistema operativo, debido a que los dos mecanismos son muy genéricos. Los principales inconvenientes de un CGI son que debe generar toda la página HTML, y que cada programa CGI que se pone en marcha lo hace en un espacio de memoria propio, por lo que si el número de usuarios que ponen en marcha un CGI a la vez es muy grande, la cantidad de recursos que ocupe ese CGI se multiplicará por el número de usuarios. Los CGI se han convertido en una alternativa relativamente obsoleta.

ASP son básicamente páginas HTML en las que se incrustan (entre marcas especiales) fragmentos de código en algún lenguaje de scripting (en general VBScript o JavaScript).

Estos fragmentos, sin embargo, se ejecutan en el servidor Web, no en el cliente. El resultado de esa ejecución es un código HTML que se inserta en lugar del código de scripting original.

El código de scripting suele usarse para acceder a bases de datos o generar información dinámica de cualquier otro tipo. ASP es una tecnología

de Microsoft, y en principio solo es soportada por *el Internet Information Server* de Microsoft.

Servlets: son programas en Java, muy similares a los applets, pero que se ejecutan en el servidor. No tienen una apariencia gráfica, a diferencia de los applets. Un servlet puede acceder a una base de datos o realizar cualquier otra tarea que permita generar contenido. Incluido en una página HTML, el servlet será ejecutado por el servidor, y su resultado se incluirá en el lugar correspondiente de la página. Una vez realizado el proceso (que requiere que el servidor Web sea capaz de tratar los servlets) al usuario le llega el HTML resultante de la sustitución.

JSP nació como una mejora de los servlets. En lugar de crear un servlet con toda su infraestructura, el programador se limita a escribir sentencias Java incrustadas en el HTML (al estilo de lo que se hace en ASP). El servidor web ejecuta ese código, y el resultado de la ejecución de esos fragmentos de código sustituye a dicho código en la página que finalmente se suministra al cliente. JSP permite acceder a bases de datos o generar otro tipo de contenido dinámico.

1.3.3 Tecnología WAP (Wireless Application Protocol)

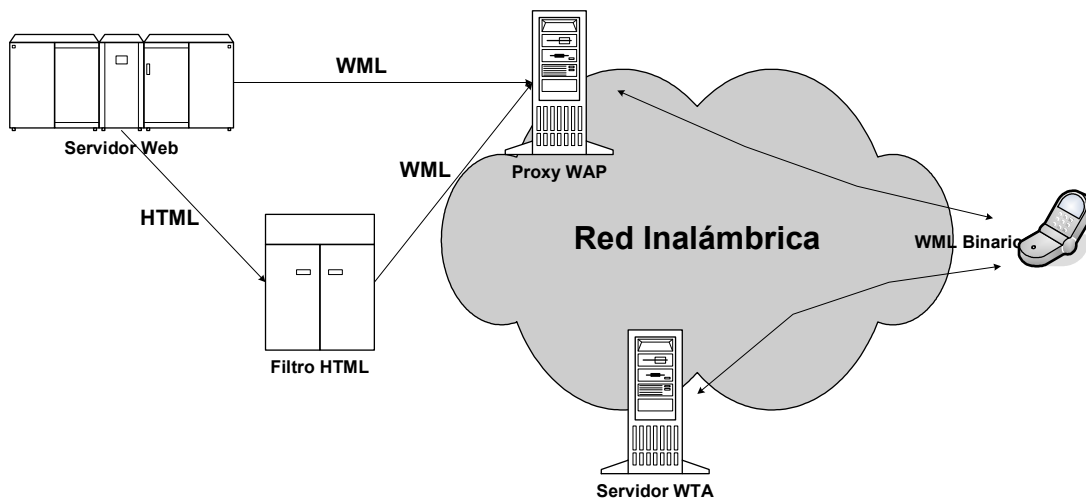
Conocido también como Internet móvil, es un protocolo para aplicaciones inalámbricas. Esta tecnología permite acceder a Internet desde dispositivos móviles.

La arquitectura WAP cuenta con un elemento llamado *Gateway*, el cual tiene como función principal permitir que los contenidos de Internet estén disponibles en el entorno móvil. Con el uso de esta tecnología, el móvil se ha convertido en una herramienta multifuncional.

El lenguaje utilizado es el WML (*Wireles Markup Lenguaje*) permite la utilización de pantallas pequeñas y navegación sin teclado. Este lenguaje de

markup está basado en el XML y es parecido al HTML; es leído e interpretado por un micronavegador instalado en el dispositivo WAP. Debido a que las capacidades de cada dispositivo varían, cada navegador puede interpretar el WML de distinta forma. Esta tecnología fue ideada por las compañías Nokia, Ericsson, Motorola y Phone.com, y el objetivo es poder ofrecer todos los contenidos y servicios de Internet a través de la red inalámbrica.

Figura 1. Una red WAP



Fuente: Servicio de Telecomunicaciones

<http://pegaso.ls.fi.upm.es/servicios/transparencias3/sld021.htm> (2003)

WAP posee una serie de características, entre las que se pueden mencionar:

- Entrega rápida y fácil de información y servicios a los usuarios de móviles

- Puede trabajar con una variedad de dispositivos inalámbricos digitales, como teléfonos móviles, radios bidireccionales, Handhelds, pagers, PDAs (Personal Digital Assistant), smartphones y comunicadores.
- Puede implantarse en cualquier sistema operativo, incluso PalmOS, EPOC, Windows CE, OS/9, JavaOS, etc.
- Es escalable, por lo que las aplicaciones pueden disponer de los recursos de red y capacidades de pantalla según lo necesiten. Los servicios son aplicables desde pantallas de una sola línea hasta terminales mucho más complejas.
- Proporciona interoperabilidad de servicio incluso entre diferentes familias de dispositivos.
- Su diseño permite que pueda trabajar con la mayoría de redes inalámbricas actuales como: CDPD, CDMA, GSM, PDC, TDMA, FLEX, iDEN, TETRA, DECT, así como para las redes futuras como UMTS.

Estas características pueden resumirse en 4 frases: base en estándares existentes, independencia de la red, independencia del dispositivo e interoperabilidad.

Entre los servicios que WAP proporciona se encuentran: navegar, realizar transacciones bancarias, reservar vuelos de avión o tren, mandar un e-mail, comprar o reservar entradas para espectáculos, consultar el tiempo y las noticias. Sin embargo, actualmente tienen serias limitaciones. Además, WAP habilita un nuevo tipo de servicios que no se encuentran en Internet, los servicios telefónicos. Algunos de estos ya están incorporados desde hace tiempo en los teléfonos móviles convencionales. Estos servicios dependen de las operadoras, que pueden ofrecerlos gratuitamente como estrategia de marketing o bajo determinadas condiciones de contrato.

Entre los principales servicios cabe destacar la administración de llamadas, el buzón de voz y la mensajería unificada (e-mail, faxes, mensajes de voz, etc.), Agenda, servicios SMS y servicios de atención al cliente.

Actualmente, las páginas y sitios Web con contenidos y servicios específicos para los usuarios de estos pequeños dispositivos surgen de manera constante, ya que muchas compañías (bancos, cadenas de televisión) han hecho alianzas con empresas de telecomunicaciones para mostrar sus contenidos en este nuevo mercado. Se espera que con el avance de la tecnología, como el diseño de avanzados micro-navegadores, menús lógicos y pantallas de mayor tamaño, el manejo de la Web a través del móvil sea más fácil.

1.3.3.1 Ventajas de la Tecnología WAP

- Portabilidad: acceso a Internet desde cualquier lugar
- Arquitectura Cliente/Servidor: dispositivos con micro-navegador y almacenamiento de servicios/aplicaciones temporalmente
- Mayor nivel de seguridad en las transacciones
- Unificación de mensajería y contenidos en servicios Internet
- Dispositivo como "*mobile wallet*" o cartera móvil en comercio electrónico
- Soporte HTTP 1.1 basado en WML (migración desde HTML)
- Soporte para la Tercera Generación de telefonía celular (aplicaciones multimedia y acceso a alta velocidad)

1.3.3.2 Desventajas actuales del mercado WAP

- Disponibilidad y coste de dispositivos adaptados en el mercado
- Limitaciones físicas de los terminales presentados (maneabilidad y usabilidad principalmente), ya que las pantallas son pequeñas,

monocromas y las posibilidades de navegación escasa, también son importantes otros factores como la CPU, la memoria y las baterías

- El coste de las llamadas podría resultar un problema, dependiendo de la compañía que preste el servicio de telefonía celular y el tipo de suscripción a la misma
- Disponibilidad de contenidos Internet con versión WAP: portales, proveedores y empresas de servicios aún muy limitadas en este protocolo
- Velocidad de transferencia baja (pico máximo de 9 Kbit/s) y no garantizada
- Alta latencia. La latencia consiste en largos períodos de tiempo (comparando con el tiempo útil de transmisión de datos) en que la sesión abierta está en estado latente, es decir, sin recibir datos

1.3.4 Audiopáginas

Para todo lo descrito anteriormente la interacción ocurre mediante un teclado y una pantalla (tanto en las páginas estáticas dinámicas como en los dispositivos móviles para WAP). Sin embargo, el ideal siempre ha sido que cada página Web pueda ser navegada a través de la voz y mediante la utilización de un teléfono convencional. Las audiopáginas cumplen con este ideal ofreciendo los contenidos Web a través de cualquier teléfono, tanto móvil como convencional.

IVR (*Interactive Voice Response*) es el software de reconocimiento de voz interactiva (IVR) habilitado para reconocimiento de voz. Permite que los clientes den órdenes hablando, en vez de oprimir teclas para acceder a la información que necesitan. Con el reconocimiento de voz los clientes pueden expresar, por medio de la voz, lo que buscan.

En el capítulo siguiente se expone en detalle la manera como funcionan las audiopáginas.

1.4 Evolución de la telefonía

1.4.1 Telefonía fija

La telefonía fija representa todavía la parte principal del mercado de telecomunicaciones en el mundo.

1.4.2 Telefonía móvil

Las telecomunicaciones han avanzado tanto que hoy en día es común que un porcentaje alto de personas posean un teléfono móvil con el cual, no importando en el lugar que estén, pueden comunicarse mediante la red inalámbrica.

1.4.3 Tecnología de tercera generación (UMTS)

UMTS "**universal mobile telecommunications system**", (Sistema de Telecomunicaciones Móviles Universal).

Es el protocolo de la tercera generación de teléfonos celulares. Está siendo desarrollado por un grupo de empresas bajo el nombre de ETSI.

Esta tecnología pretende realizar la transmisión de datos y voz, teniendo abiertos dos canales a la vez y llegando a alcanzar velocidades de hasta 2 Megabits por segundo (hasta doscientas veces la capacidad actual de cualquier teléfono móvil; sin embargo, esta solamente se alcanzará en el marco de una adecuada infraestructura de redes.)

El terminal UMTS pasará de ser un simple dispositivo, útil solamente para realizar conversaciones telefónicas, a convertirse en un potente dispositivo de comunicaciones con capacidades avanzadas, de imagen y sonido, como videotelefonía o videoconferencia, acceso a información similar a una computadora personal y características enfocadas a la facilidad de uso, como personalización de portales, servicios personalizados y servicios basados en la localización puntual del usuario.

Esta tecnología hará que la conexión a Internet sea fluida y con capacidad para transmitir imágenes y sonidos.

1.4.4 Telefonía e Internet

Estos son dos sectores de gran expansión dentro del mundo de las telecomunicaciones. Al fusionarse han dado la posibilidad de estar libres del ordenador sin desconectarse de la WWW y permitir, de esta manera, que la red no solo se mire sino se oiga.

1.5 Voz sobre IP (Voice Over IP, VoIP)

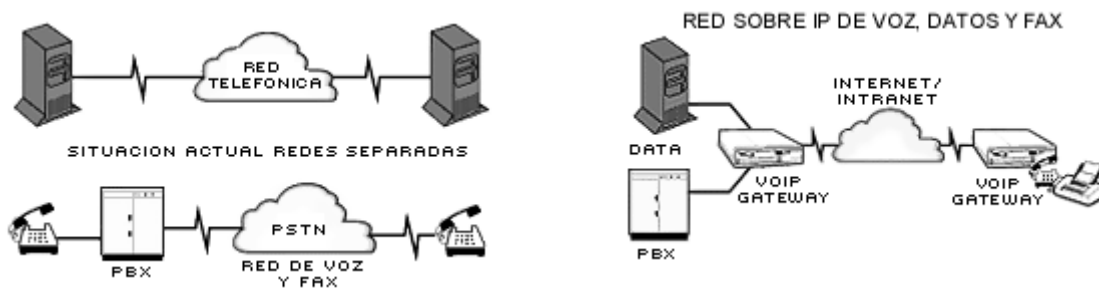
Desde los años 80 la red telefónica funciona de manera igual. Sin embargo, en las redes de datos los avances han sido muy grandes e importantes tanto en fiabilidad, en capacidad como en costos. Todos estos adelantos se pueden empezar a aplicar a las comunicaciones de voz gracias a los últimos desarrollos presentados sobre la tecnología Voz IP.

La voz sobre IP consiste en convertir las señales de voz en paquetes de datos, los cuales son transportados de forma comprimida en lugar de utilizar las tradicionales líneas telefónicas se utilizan redes de datos.

La transmisión basada en paquetes ha dejado atrás a la transmisión conmutada por circuitos, tomando el tráfico de la red pública telefónica y

colocándolo en redes IP. Esta forma de transmisión permite que las señales de voz se encapsulen en paquetes IP tanto nativo como IP por Ethernet, Frame Relay, ATM o SONET.

Figura 2. Redes de telefonía actual vrs Red sobre IP



Fuente: Voz sobre IP Corporativa.

http://www.ebosa.cl/pags/soluciones/redes_wan/cont2_voz_ip.html (2003)

La especificación H.323, aprobada en 1996 por el ITU (International Telecommunications Union) y revisada en enero de 1998, es la base de las arquitecturas interoperables de voz sobre IP. Sugiere la manera de establecer, enrutar y terminar llamadas telefónicas a través de Internet. Proporciona el estándar necesario para que la evolución de la voz sobre IP sea común entre los diversos fabricantes.

Este estándar define tres elementos:

- Terminales. Son el sustituto de los actuales teléfonos lo cual permite aumentar la funcionalidad y sencillez de uso, tanto mediante hardware como mediante software.
- Gatekeepers. Sustituyen a las actuales centralitas telefónicas, por lo que se convierten en el centro de toda la organización de Voz IP.

- Gateways. Interfaces de telefonía con la red. Su misión es la de enlazar la red VoIP con la red telefónica analógica o RDSI, por lo que el enlace se puede realizar de forma totalmente transparente para el usuario. Se puede considerar al Gateway como una caja que, por un lado, tiene un interfase LAN y por el otro dispone de uno o varios de los siguientes interfaces:
- FXO. Para conexión a extensiones de centralitas o a la red telefónica básica.
- FXS. Para conexión a enlaces de centralitas o a teléfonos analógicos.
- E&M. Para conexión específica a centralitas.
- BRI. Acceso básico RDSI (2B+D)
- PRI. Acceso primario RDSI (30B+D)
- G703/G.704. (E&M digital) Conexión específica a centralitas a 2 Mbps.

Como ya se mencionó, entre los objetivos de este estándar está la interoperabilidad entre equipos de diferentes fabricantes, para lo cual se han fijado aspectos como: codificación de la voz, supresión de silencios, direccionamiento y se han establecido nuevos elementos (servicios de directorio, y transmisión de señalización por tonos multifrecuencia) para permitir la conectividad con la infraestructura telefónica tradicional.

El estándar H.323 para el VoIP se apoya en una serie de protocolos que cubren los distintos aspectos para la comunicación:

- Direccionamiento:
 1. RAS (Registration, Admission and Status). Protocolo de comunicaciones que permite a una estación H.323 localizar otra estación H.323 a través de el Gatekeeper

2. DNS (Domain Name Service). Servicio de resolución de nombres en direcciones IP con el mismo fin que el protocolo RAS, pero a través de un servidor DNS

Señalización:

1. Q.931 Señalización inicial de llamada
2. H.225 Control de llamada: señalización, registro, admisión y “paquetización” / sincronización del stream (flujo) de voz
3. H.245 Protocolo de control para especificar mensajes de apertura y cierre de canales para streams de voz

- Compresión de voz:

1. Requeridos: G.711 y G.723
2. Opcionales: G.728, G.729 y G.722

- Transmisión de voz:

1. UDP. La transmisión se realiza sobre paquetes UDP, pues aunque UDP no ofrece integridad en los datos, el aprovechamiento del ancho de banda es mayor que con TCP.
2. RTP (Real Time Protocol). Maneja los aspectos relativos a la temporización, marcando los paquetes UDP con la información necesaria para la correcta entrega de los mismos en recepción.

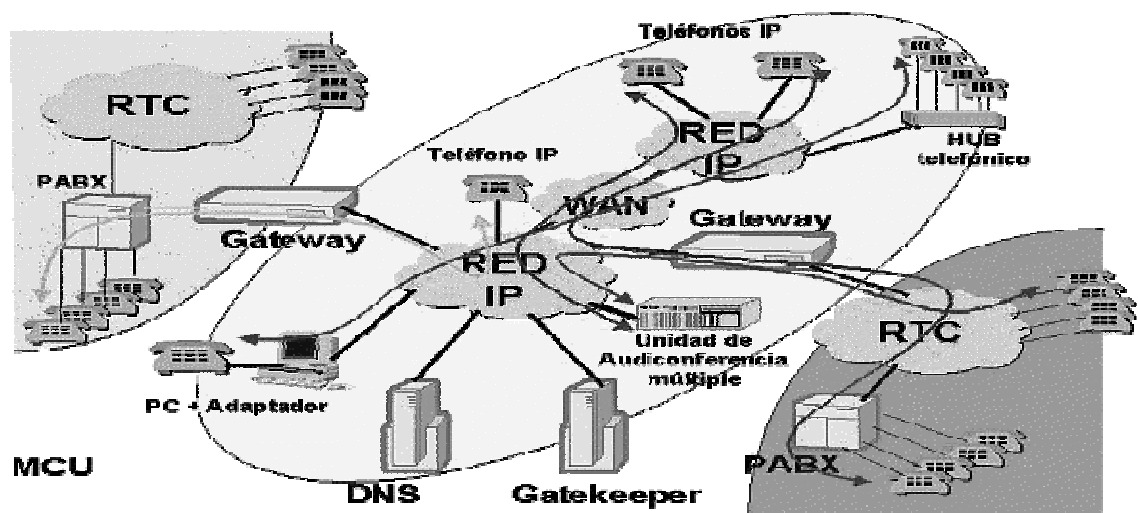
- Control de la transmisión:

1. RTCP (Real Time Control Protocol). Se utiliza principalmente para detectar situaciones de congestión de la red y tomar, en su caso, acciones correctoras.

Actualmente es posible utilizar varios elementos ya disponibles en el mercado y que según diferentes diseños permiten construir las aplicaciones VoIP. Algunos de estos elementos son:

- Teléfonos IP
- Adaptadores para PC
- Hubs Telefónicos
- Gateways (pasarelas RTC / IP)
- Gatekeeper
- Unidades de audioconferencia múltiple. (MCU Voz)
- Servicios de directorio

Figura 3. Elementos de una red VoIP



Fuente: El estándar VoIP.

<http://www.comunicaciones.unitronics.es/tecnologia/voip.htm> (2003)

En conclusión, VoIP es la tecnología que en poco tiempo muchas empresas comenzarán a implantar, ya que les permitirá aprovechar mejor la red de datos y, por otro lado, minimizar los costos ocasionados por la red telefónica. Y todo esto gracias a que los teléfonos IP ya están disponibles y los principales operadores mundiales están promoviendo activamente el servicio IP a las empresas, ofreciendo calidad de voz a través del mismo.

1.6 Tecnología de voz

1.4.1 Ingeniería lingüística

La lengua es un factor determinante en las comunicaciones directas entre seres humanos sin embargo, al tratar de aplicarla a la comunicación con los sistemas, servicios y aplicaciones utilizados diariamente, este factor se convierte en un obstáculo. Por tal motivo, existe la ingeniería lingüística, la cual tiene como objetivo utilizar los conocimientos que el ser humano tiene de la lengua para mejorar la interacción con los sistemas informáticos, mediante un análisis, selección y presentación de la información de manera más eficaz, así como proveyendo medios para generar lenguaje natural y de traducción. El ideal de todo esto es que las máquinas sean capaces de comprender el lenguaje humano, traducir a los diversos idiomas y poder generar resultados orales e impresos, lo cual permitirá que la máquina se convierta en la mejor ayuda para el ser humano al momento de entender mejor y más rápido a sus semejantes y pueda cooperar con mayor eficacia en aspectos económicos, políticos y de cualquier otra índole.

Cuando la computadora sea capaz de reconocer, comprender y generar el habla, producirá la interacción más eficaz entre las personas y los ordenadores. Esta interacción permitirá superar el problema del exceso de información, ya que se obtendrá únicamente lo necesario, cumpliendo con la precisión y sensibilidad deseadas para satisfacer las necesidades de información.

Todos los avances que se han conseguido mediante la ingeniería lingüística darán las siguientes ventajas:

- Poder ahorrar tiempo en la búsqueda de información ya que se accederá a la misma de manera más pronta y sin tener sobrecargas, pues se encontrará solamente lo necesitado

- Posibilidad de realizar una comunicación directa (como si se estuviera hablando con cualquier persona) con la computadora, se encuentre ésta en cualquier lugar (casa, trabajo, carro, centro comercial, etc.), solicitándole cualquier tipo de información como saber que sucede a nuestro alrededor, tanto nacional como internacionalmente
- Hacer negocios seguros mediante el teléfono, interactuando directamente y de manera fiable con sistemas informáticos que funcionan vocalmente

Existen 4 elementos fundamentales de la ingeniería lingüística:

a. Identificación de la voz

Como se sabe, la voz humana es única para cada persona, lo que permite identificar a alguien en particular cuando está hablando. Este reconocimiento es utilizado en algunas ocasiones para acceder a un servicio o recurso. Sin embargo, debido a diversos factores (imitadores, ruido, enfermedad, entre otros) no se ha podido concretar como un medio confiable para la autorización.

b. Reconocimiento del habla

La voz es recibida mediante ondas analógicas que la computadora se encarga de analizar y poder, de esta forma, obtener los fonemas (unidades de sonido) que son los que conforman las palabras. A pesar de los numerosos avances de la tecnología en cuanto al tratamiento de la voz, aún existen muchos problemas para poder utilizarla para la comunicación con el ordenador, entre los que se pueden mencionar:

La capacidad de reconocer el habla continúa en lugar de reconocer palabras independientes separadas por una pausa.

- Reconocer a cualquier hablante (lo cual contrasta con el entrenamiento que se le da al sistema para que reconozca una voz específica)
- La interferencia del ruido, el cual proviene del entorno donde se encuentra la persona o del medio de transmisión.
- El problema de los acentos y dialectos del lenguaje hablado, que casi nunca corresponden a las normas gramaticales de la lengua escrita.

c. Comprensión del lenguaje natural

La comprensión del lenguaje se da mediante una serie de etapas. En la primera etapa se logra una comprensión regularmente parcial, lo que permite conseguir una clasificación inicial de textos. Este análisis superficial da lugar a la segunda etapa, en la cual ya es posible seleccionar las partes interesantes de un texto. En la tercera etapa se realiza un análisis semántico que da como resultado la comprensión del contenido dentro de cierto marco establecido. Luego de estas tres etapas se obtienen modelos semánticos que son utilizados para representar significados de determinada lengua.

d. Generación del habla

Para poder generar el habla se utilizan plantillas, las cuales son rellenas a través de la reproducción de grabaciones o concatenando fonemas. En esta generación es importante tomar en cuenta aspectos como intensidad, duración y acento, para que la respuesta producida sea lo mas cercano a lo natural. Si se desea establecer un diálogo se necesita básicamente de todos los elementos antes mencionados: identificación, reconocimiento, comprensión y generación del habla. Acoplando estos elementos a una herramienta gráfica para estructurar aplicaciones se logra crear sistemas que se encarguen de la gestión automatizada de las llamadas telefónicas, entre otros aspectos.

1.4.2 Accesibilidad y participación

Todas las comunicaciones que se den mediante la voz ampliarán la utilidad de los sistemas y servicios, ya que todo tipo de personas podrá hacer uso de los mismos.

Los sistemas que tengan una interacción con sus usuarios mediante el lenguaje humano desde cualquier punto de acceso (casa, trabajo o lugar público) serán los que predominarán en poco tiempo y, por ende, los más usados.

“La aplicación de las tecnologías lingüísticas ayudará a muchas personas cuyas vidas se ven afectadas por algún tipo de discapacidad. Los ordenadores dotados de sistemas de comprensión de las lenguas, capaces de escuchar, ver y hablar, ofrecerán nuevas oportunidades para acceder a los servicios desde el hogar y participar en el lugar de trabajo”¹.

Como se observó en la descripción de estos últimos párrafos la aplicación de la ingeniería lingüística provee muchos beneficios, a continuación se enumeran los más sobresalientes:

- Mejores servicios de las administraciones públicas y los organismos que prestan servicios públicos
- Mayores posibilidades de acceso a la información, merced a una utilización más sencilla de los ordenadores y de los servicios de información
- Ahorro de tiempo gracias a la utilización de ordenadores inteligentes como agentes nuestros
- Mejora de la calidad de la información registrada en los sistemas de información
- Menor tensión en las situaciones en que tanto las manos como la vista estén ocupadas

- Mayores oportunidades para integrar a los discapacitados en el trabajo y en las actividades sociales cotidianas

Cuando se establece una comunicación entre personas, básicamente se distinguen 2 partes: la emisión de la información y la recepción de la misma. De manera similar, en las tecnologías de voz se distinguen estas 2 partes:

1. la lectura que el ordenador realice de un texto de manera similar a como lo haría una persona (conversión texto en habla)
2. la posibilidad de convertir automáticamente lo expuesto mediante la voz en un texto escrito (reconocimiento de habla)

1.4.3 El reconocedor de voz natural

La acción que ejecute una persona al hablar con el ordenador como si lo hiciera con otra persona, y, a la vez, que el ordenador sea capaz de reconocer y comprender lo que se le ha dicho, eliminando la rigidez tradicional en la interacción hombre/máquina, es la función del reconocedor de voz.

Hoy en día las computadoras automatizan satisfactoriamente muchas tareas que para el ser humano son difíciles, repetitivas o requieren mucho tiempo. Sin embargo, les es un poco más difícil realizar tareas “aparentemente sencillas”, como reconocer la voz, ya que para efectuar este reconocimiento deben realizar una serie de tareas: ordenación, búsqueda de información y realización de cálculos matemáticos.

3.4.1.1 Arquitectura del reconocedor de habla natural

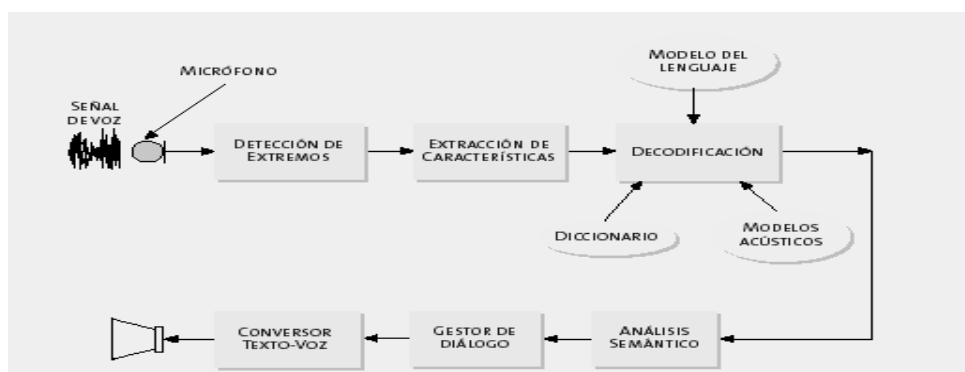
El procedimiento para llevar a cabo el reconocimiento de voz involucra 4 tipos de información:

1. Modelos acústicos: tiene que ver con la identificación de sonidos por parte del reconocedor.

2. El diccionario: indica qué sonidos forman cada palabra.
3. Modelo del lenguaje: se encarga de proporcionar la información para formar las frases mediante la combinación de palabras.
4. Los sistemas conversacionales: este es el más complejo, ya que el reconocedor debe conocer o predecir la siguiente frase que el locutor le dirá.

El proceso de dialogo con la maquina es el siguiente: la voz se recibe y se convierte en una señal eléctrica analógica, luego pasa a ser digitalizada, en seguida pasa al detector de extremos (el cual detecta la presencia de voz). Posteriormente el extractor de características es el encargado de tomar la información relevante para el reconocimiento. Esta información es pasada al decodificador, el cual usa los modelos acústicos y del lenguaje junto al diccionario, para poder generar la frase reconocida. Los últimos pasos son efectuados por el analizador semántico y el gestor de diálogo, el primero para extraer el significado de la frase y el segundo para tomar la decisión adecuada dependiendo de la situación.

Figura 4. Diagrama a bloques de un sistema de dialogo



Fuente: Estado del arte en tecnologías de Voz.

<http://www.tid.es/presencia/publicaciones/comsid/esp/20/8XX.PDF> (2003)

3.4.1.2 Características que afectan al reconocedor de habla natural

Independencia del locutor

Para el ser humano es fácil identificar a una persona por su voz y, al mismo tiempo, le es posible obtener información sobre el interlocutor. Pero en el caso de los sistemas de reconocimiento, las diferencias entre voces tienen efectos negativos en la tasa de aciertos.

Efectos del canal y del ruido

En diversos experimentos realizados se ha comprobado que el ruido de fondo afecta relativamente poco a las tasas de reconocimiento del ser humano, ya que el ser humano tiene la capacidad de identificar la fuentes de sonido y su separación debido al proceso realizado posteriormente por el cerebro, así como por la capacidad de predicción del cerebro gracias a la serie de fuentes de conocimiento que tiene y a la capacidad de adaptación tanto a ruido como a la cancelación de este.

Tales factores demuestran que los sistemas de reconocimiento deberían utilizar más información para poder realizar mejor su trabajo.

Características del habla espontánea

El habla espontánea es la que día a día el ser humano usa para comunicarse con sus semejantes. Sin embargo, este hecho presenta algunas dificultades y, a la vez, retos para la tecnología del habla; por ejemplo:

1. Existen muchos sonidos que desaparecen o que pierden parte de sus propiedades acústicas.
2. Hay sonidos que no pertenecen al idioma "normativo", pero sí pertenecen a un determinado dialecto del mismo.
3. Los modelos del lenguaje funcionan peor que el estilo de habla leído, ya que reducen la capacidad de corrección del modelo del lenguaje.

La solución para estas dificultades será el resultado de la sinergia de múltiples áreas de conocimiento y ordenadores más potentes

1.4.4 El conversor texto a voz

1.4.4.1 Sistemas de conversión de texto a voz (tts)

Básicamente se distinguen dos bloques funcionales:

1. El bloque de proceso lingüístico

el cual se encarga de obtener la cadena de sonidos a pronunciar, así como la información sobre como hacerlo y su correcta entonación.

2. El bloque de síntesis de voz

el cual concatena unidades acústicas que han sido diseñadas y grabadas, con lo cual se obtiene la síntesis de voz.

1.4.4.2 La conversión texto-voz avanzada

Este tipo de conversión tiene incorporadas varias características, las cuales permiten su integración en la información manejada hoy en día: corrección automática de texto del correo electrónico, deletreo y lectura adaptada de términos en Internet, detección automática del idioma de un texto a lectura de documentos ofimáticos (word, excel, etc.)

1.4.4.3 Aplicaciones

Existen muchas aplicaciones de la conversión de texto a voz, entre las que se pueden citar: recepción de correo electrónico a través del teléfono, lectura de un texto, información metereológica, estado del tráfico, horarios de transportes públicos, farmacias de guardia, consulta de carteleras de cine, etc. En el caso de estas últimas, la información que se da a conocer debe ser actualizada constantemente, lo que requeriría de constantes grabaciones para poder presentarlas. Sin embargo, al utilizar este tipo de conversión las actualizaciones se hacen con mayor facilidad y no es necesario tener

operadores contratados para compartir este tipo de información, ya que toda se encuentra almacenada en una base de datos.

1.7 El ruido, amenaza latente para los sistemas de reconocimiento de voz

El ruido es cualquier señal acústica o eléctrica que contamine la señal de voz que se desea reconocer. El ruido se puede clasificar en 3 grupos:

1. *Ruido estacionario* se mantiene constante en un período de tiempo, como: tubos fluorescentes, ventiladores de los ordenadores, etc.
2. *Ruido no estacionario* aquí entran todos los ruidos impredecibles, los cuales aparecen de forma intermitente, la mayoría de ruidos existentes son de este tipo.
3. *Voces de fondo* Son consideradas como ruido no estacionario; su efecto es dañino en los sistemas de reconocimiento, ya que degradan las tasas de reconocimiento, al contaminar la voz. Además, un servicio puede llegar a fallar por reconocer la voz de fondo.

Cuando el reconocimiento se ejecuta mediante la línea telefónica, el problema del ruido es aún más grave, ya que los diversos circuitos eléctricos y electrónicos por donde pasa la voz antes de llegar al sistema producen variados efectos en la señal de voz, por lo que se convierten en enemigos del reconocedor de voz.

A similar problema se enfrenta la transmisión de voz mediante las redes de datos, ya que los paquetes IP son de longitud variable y el tráfico de datos suele ser a ráfagas. Sin embargo, con la implementación del protocolo RSVP se logra dar prioridad a los paquetes de voz.

2. INTRODUCCION A LOS PORTALES DE VOZ

El termino portal es un sinónimo de puerto, que por lo regular se utiliza para nombrar un sitio Web que sirva como punto de partida para iniciar la actividad de navegación en Internet.

Definición de un portal de voz

Al escuchar hablar de Internet, inmediatamente la asociación que se hace es con un ordenador y un ISP (*Internet Service Provider*). Con la incorporación de la tecnología WAP, esa asociación llega a abarcar el acceso mediante teléfonos móviles que permitan el uso de este protocolo. Sin embargo, hoy en día es importante conocer que existe una alternativa más a estos 2 medios de acceso; su nombre es: "Portal de voz", conocido también como: "*Voice Browser*".

A continuación se presentan varias frases que dan a conocer lo que es un portal de voz:

- "Los portales de voz ponen al alcance del oído Internet"².
- "Un Portal de Voz no es más que una tecnología basada en el reconocimiento de voz"

- “Los Portales de voz permiten acceder a información alojada en servidores de la empresa tanto en intranets como en Internet, de la manera más natural para el usuario: hablando a través de cualquier teléfono, esta facilidad de operación se logra con soluciones que interconectan las redes telefónicas con dichos servidores y combinan las más avanzadas tecnologías de reconocimiento y síntesis vocal”³.
- “Un portal de voz es el que permite acceder a Internet mediante la voz. Ofreciendo los contenidos que actualmente residen en la Web a través de la interfaz más accesible que se conoce: el teléfono”.

Con este nuevo medio de acceso, en vez de utilizar navegadores se usan teléfonos (de cualquier tipo); en lugar de utilizar una dirección Web se usa un número de teléfono, y para ir accediendo a la información que se desea, solo es necesario usar la voz.

Por tal motivo los portales de voz se convierten en un nuevo canal de interacción basado en comunicar mediante el teléfono fijo, móvil, o teléfono público los contenidos Web.

Los portales de voz constituyen la mejor alternativa para captar nuevos usuarios que no cuentan con una computadora personal, como también para todos aquellos usuarios que viajan continuamente y que se complacerían de conectarse a Internet a través de un teléfono, en lugar de usar una PC.

Evolución histórica

Durante varios años para poder presentar contenidos Web se utilizó el html. Luego cuando se comenzó a pensar en la posibilidad de proveer contenidos Web para que fueran consultados mediante dispositivos móviles, se creó el WML, y como un paso mas se creó el *VoiceXML*, el cual permite crear contenidos Web para ser accedidos mediante la voz y, al igual que WAP, utilizando una plataforma telefónica.

Esto se hizo con el objetivo de extender los conceptos tradicionales de ISP, portales, buscadores, navegadores, etc., de Internet Tradicional. Esta incorporación del teléfono convencional al mundo de Internet ha abierto la posibilidad de nuevas forma de negocio por teléfono, así como publicidad y comercio electrónicos. En la figura 5 se puede observar de una mejor forma la evolución histórica del acceso a contenidos Web.

Entrando un poco más a fondo en la historia de los portales de Voz, se sabe que los mismos surgieron a mediados de 1999 en Estados Unidos. La iniciativa fue de un grupo de ex trabajadores de *Netscape*, quienes querían hacer posible el acceso a Internet desde cualquier parte, según información publicada, se sabe que el éxito del primer portal de voz, llamado "*TellMe*", fue tan grande que el servicio colapsó el primer día de su funcionamiento debido a la fuerte demanda de llamadas que recibió.

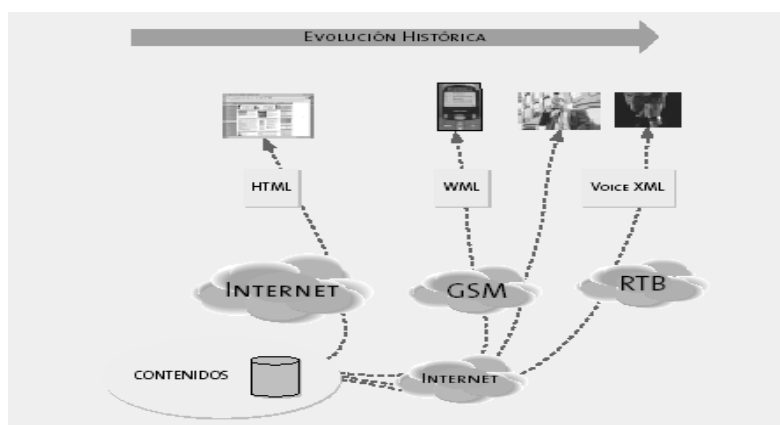
Los usuarios vieron en esta opción una forma de acceso sencilla, con lo que podían prescindir de un ordenador. Es así como la tecnología, en forma de programas de reconocimiento de voz y de conversión de texto a voz, ha sido fundamental para la creación de este nuevo servicio.

Se puede decir que acceder a Internet con solo utilizar la voz es el método más natural para el ser humano. No existe nada más sencillo que el hecho de poder navegar por Internet interactuando mediante órdenes vocales, es decir, como si se estuviera conversando con otra persona.

El portal de voz es mucho más que una navegación vocal, ya que pretende relacionar el mundo de Internet con el mundo de la voz, utilizando para ello un conjunto de servicios que amplíen, complementen y den un valor añadido al acceso vocal a Internet. Esta es la razón por la cual se incluyen servicios como el e-mail por voz, las llamadas telefónicas y los envíos de *e-mails* a direcciones de una agenda Web, la búsqueda de empresas y personas de forma vocal y, por supuesto, el acceso a contenidos específicos Web. Etc.

Este nuevo medio de acceso a la Web es de sumo interés principalmente para las operadoras telefónicas, ya que puede suponer una “reinención de la telefonía” para otras potencialidades de los teléfonos que aún no se han explorado. Así mismo, como se mencionaba anteriormente, esto permitirá realizar actividades y negocios asociados a la publicidad y el comercio electrónico, y todo mediante el teléfono.

Figura 5. Evolución histórica del acceso a Internet



Fuente: Estado del arte en tecnologías de Voz.

<http://www.tid.es/presencia/publicaciones/comsid/esp/20/8XX.PDF> (2003)

Servicios de un portal de voz

Ha sido común que los servicios telefónicos tradicionales proporcionen a las personas información sobre noticias, tráfico, eventos de diversa índole, así como información meteorológica, etc. Pero para poder brindar todos estos servicios el costo ha sido muy grande, ya sea porque es ofrecido mediante operadoras o porque la información debe ser actualizada constantemente, para lo cual se requiere de grabaciones.

Con la implementación de los portales de voz, es posible realizar las actualizaciones de manera rápida, y el costo asociado es mucho menor que el de una grabación o el de la operadora.

Existe una diversidad de servicios que se pueden brindar mediante un portal de voz, entre los que destacan:

- Navegación por Internet.
- Noticias: nacionales, internacionales, bursátiles, meteorológicas, etc.
- De localización geográfica: restaurantes, hoteles, farmacias cercanas, etc.
- Mensajería Unificada
- Notificaciones a móviles
- Envío y recepción de *e-mail*
- Agendas vocales:
 - Personal
 - De negocios
- Directorios de empresa
- Servicios de contenidos
- Entretenimiento: horóscopos, chistes, conversaciones con consejeros virtuales, etc.
- Servicio de directorio (páginas blancas)

A continuación se da una breve descripción de las funcionalidades que pueden incluir algunos de los servicios antes mencionados.

2.3.1 Envío y recepción de e-mail

Para este servicio se le puede brindar al usuario la opción de configurar su perfil para el acceso a su cuenta de correo electrónico normal.

Cuando el usuario se conecta telefónicamente, mediante la conversión de texto a voz se le da a conocer el contenido de su buzón de correo. Así mismo, se le podría proporcionar un conjunto de funcionalidades como:

- Elección del idioma de atención del servicio telefónico
- Detección del idioma de los correos electrónicos
- Lectura de los correos en el idioma detectado
- Corrección del texto del mensaje, esto es usado para evitar en lo posible los errores de escritura comunes en los correos electrónicos (signos de puntuación, acentuación, etc.)
- Respuestas a los correos del buzón del usuario, usando un anexo WAV grabado por el usuario
- Envío y reenvío de correos electrónicos a direcciones de la agenda del usuario utilizando siempre anexos WAV grabados previamente por el usuario
- Lectura de anexos ofimáticos: word, excel, power point, etc.

2.3.2 Agendas vocales

Al igual que con el correo electrónico, el servicio de las agendas personales y de negocios puede ser personalizado por los usuarios tanto vía Web o voz.

Las opciones que podrían tener para este tipo de servicio son: dar de alta, borrar y modificar vocalmente las direcciones de las agendas. Si existiera información que no se puede llenar vía voz, posteriormente el usuario podría hacerlo vía Web.

Aparte de poder personalizar este servicio, el usuario podría enviar correos electrónicos o hacer llamadas a las direcciones que haya configurado. Para poder asegurar la autenticación del usuario y posteriormente realizar el cobro respectivo, este servicio se auxiliaría del verificador del locutor, el cual comprueba que el usuario es quien dice ser (mediante un certificado vocal grabado previamente por el usuario). En algunas ocasiones, además de utilizar este verificador, también se podría solicitar un PIN, con lo que se puede lograr mayor fiabilidad de la validación.

2.3.3 Directorio de empresas

Este tipo de servicio está enfocado a las diversas empresas que deseen aparecer en el directorio y permitir de esta forma que los usuarios puedan contactarlos vía correo electrónico o mediante una llamada telefónica, para lo cual deben ser dadas de alta por el operador del servicio correspondiente. El uso de este servicio por parte del usuario es realmente sencillo; únicamente debe pronunciar el nombre de la empresa, y el sistema le preguntará si desea enviar un correo electrónico o hacer una llamada. Al igual que el servicio anteriormente descrito, el sistema debe autenticar al usuario; de lo contrario sería un problema grave el cobro. Este tipo de servicio también podría permitir a todos los usuarios del directorio tener la opción *Follow on* (llamadas concatenadas).

2.3.4 Servicios de contenidos

Existen varios servicios que también pueden ser ofrecidos mediante un portal de voz entre los que se pueden describir:

- Servicios de noticias: incluye diversidad de información actualizada, sobre acontecimientos de diversa índole
- Información bursátil: lo que incluye información sobre los mercados financieros, cotizaciones, índices, etc.
- Información horaria: hora en punto, hora de otros países, despertador, etc.
- Localización: información relacionada con la posición del usuario, para proporcionarle informaciones sobre viajes, gastronomía, ocio, etc.
- Entretenimiento: chistes, loterías, etc.

Escenarios de operación para un portal de voz

Para poder prestar el servicio de portal de voz, es necesario contar con una plataforma la cual constituye el elemento intermediario entre el usuario que realiza la llamada y el servicio, y se encarga de suministrar los servicios de usuario y de red que requieren recursos especializados (voz, datos, etc.) y/o recursos de comunicaciones (operadoras, protocolos entre otros).

En la figura 6 se pueden observar 2 entornos de operación: por RTB y por voz sobre IP y 3 escenarios.

En el primer escenario se puede observar una llamada de forma tradicional es decir por RTB, la cual es direccionada hacia la plataforma de acceso vocal para luego acceder al servicio de portal de voz.

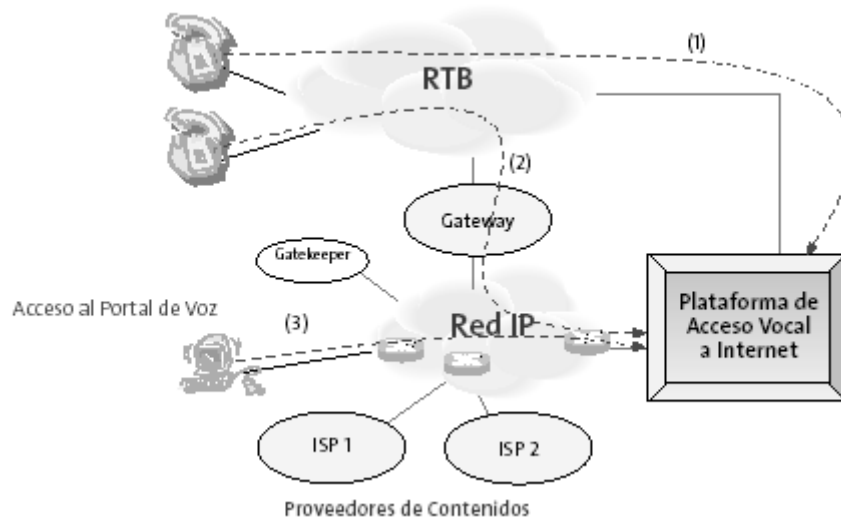
En el segundo escenario el primer recorrido de la llamada es por RTB pero luego pasa a ser transformada por el *gateway en voz sobre IP*, y a través de VoIP se transmite hacia la plataforma de acceso vocal (que en este caso está conectada directamente a la red IP).

En el tercer escenario, la llamada se realiza desde una PC, la cual desde su origen ya tiene formato de VoIP, por lo cual accede hacia la plataforma mediante la Red IP conectada.

Como se puede observar en los 3 casos, el servicio es completamente ajeno al medio de acceso de la llamada, ya que en los tres se comporta de forma similar.

Si llegaran a existir problemas de carácter regulatorio al momento de prestar el servicio, se podría contemplar la introducción de discriminadores de voz/datos en los propios servidores de los terminales, con lo que se tendría el mismo número de red tanto para acceder a los datos mediante PC, como para el acceso mediante voz.

Figura 6. Escenarios de operación



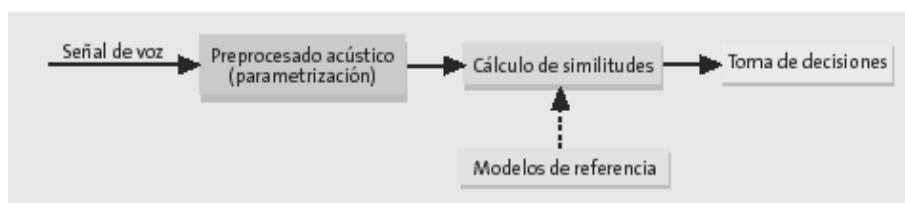
Fuente: Portales de Voz: Internet en el teléfono
http://www.tid.es/presencia/publicaciones/comsid/esp/19/ART_2.PDF (2003)

Componentes de un portal de voz

Para poder poner a funcionar un portal de voz se deben considerar todos los componentes necesarios, entre los que se encuentran:

- Un reconocedor de habla que se encarga de recoger las frases pronunciadas por el usuario. En la figura 7 se observa un modelo aplicable al reconocedor de habla.

Figura 7. Reconocedor de habla



Fuente: Tecnología de voz

<http://www.tid.es/presencia/publicaciones/comsid/esp/24/art2.pdf> (2003)

- Un conversor de texto a habla que se encarga de proporcionar las respuestas al usuario. Para este componente se cuenta con la ayuda de la tecnología lingüística (expuesta en el capítulo anterior). La misma hace posible acceder a través de la voz a los contenidos de cualquier texto escrito que se encuentre en formato digital y que es necesario en la recepción de información y en la realización de transacciones.
- Un sistema de comprensión, el cual es necesario para interpretar las peticiones del usuario.
- Un gestor del diálogo, el cual se encarga de organizar la “conversación”

Estos componentes dan vida al *browser* vocal o navegador de voz, ya que le dan la capacidad de reconocer la voz y de contestar oralmente a los requerimientos de los navegantes y, de acuerdo al navegador de que se

trate, pueden utilizar diferentes plataformas: teléfonos comunes, celulares, televisores digitales, minicomputadoras, *handhelds* y *palmtops* con conexión celular a la Web.

Arquitectura de un portal de voz

A continuación se describen los servicios que, al unirse, permiten acceder a los contenidos Web mediante el teléfono.

- 1) Medio de acceso (RTB o VoIP)
- 2) Plataforma de Acceso Vocal
 - Browser Vocal
- 3) Contenidos VoiceXML

El medio de acceso es el que interviene en el primer tramo de la llamada que, como se explicaba en los escenarios de operación se puede realizar por RTB o por VoIP, para, luego, transmitir a la plataforma de acceso vocal.

La plataforma de acceso vocal a Internet es el servicio que permite las interacciones automáticas con los usuarios por medio de conversión texto a voz, reconocimiento de voz y DTMF (detección de tonos multifrecuencia). Esta plataforma mediante el browser vocal o navegador de voz se encarga de buscar los contenidos VoiceXML solicitados para presentárselos al usuario.

De tal forma, el usuario puede interactuar hablando de forma natural y solicitando información muy diferente, mientras que el sistema responde con voz sintetizada. Por lo regular los portales de voz están hechos para reconocer palabras clave como “noticias”, “restaurantes”, “entretenimiento”, etc. Y de esta forma se va accediendo a la información requerida.

Detalle de la plataforma de acceso vocal

Existen varios elementos que interactúan y/o conforman la plataforma de acceso vocal.

A continuación se describen varios bloques funcionales en los que se puede dividir la plataforma de acceso vocal para la prestación del servicio de voz

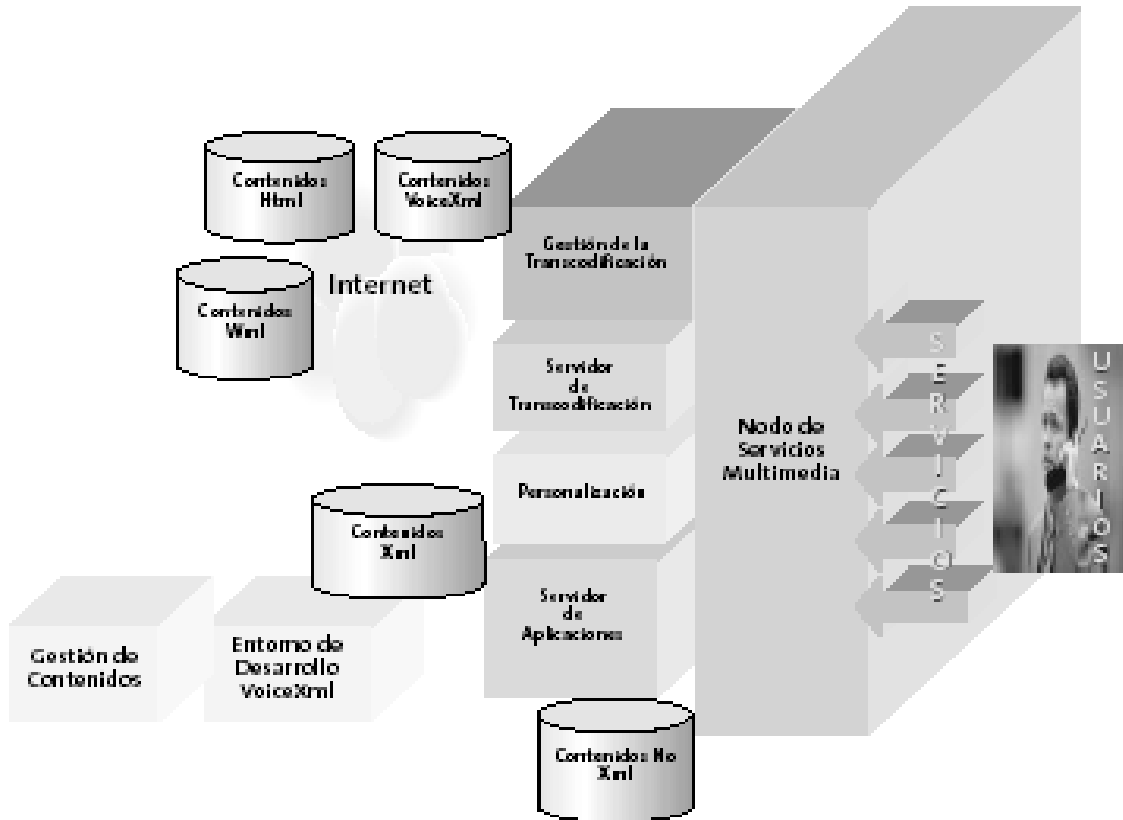
- El nodo de servicios multimedia

Este elemento presta múltiples servicios. Sin embargo, de manera principal se encarga de buscar los contenidos VoiceXML solicitados durante la navegación vocal para presentárselos al usuario.

Se trata del enlace de unión entre el mundo de la voz y el mundo IP; además, proporciona todo lo que se necesite para el tratamiento de llamadas salientes, tarificación, autenticación, validación, etc.

Para la prestación del servicio de acceso vocal a Internet, en función de que los contenidos sean propietarios o de Internet, y del formato de los mismos (No XML, html, wml o VoiceXML), la plataforma se apoyará, para interpretarlos, en alguno/s de los bloques funcionales que se describen a continuación.

Figura 8 Plataforma de acceso vocal a Internet



Fuente: Portales de Voz: Internet en el teléfono
http://www.tid.es/presencia/publicaciones/comsid/esp/19/ART_2.PDF (2003)

- El Servidor de Aplicaciones

Proporciona el acceso a información de cualquier origen y de ella extrae los datos que necesita el servicio que esté ofreciendo la plataforma. El servidor de aplicaciones podrá estar distribuido físicamente en tantas máquinas como sea necesario para la prestación eficaz del servicio.

Como se puede observar en la figura 9, el servidor de aplicaciones se comunica con cada uno de los siguientes componentes del servicio:

El Servidor de Correo Electrónico, al que accede para lectura y envío de correo.

El Servidor de Contenidos, mediante el que obtiene los contenidos propietarios que se ofrecen en el servicio.

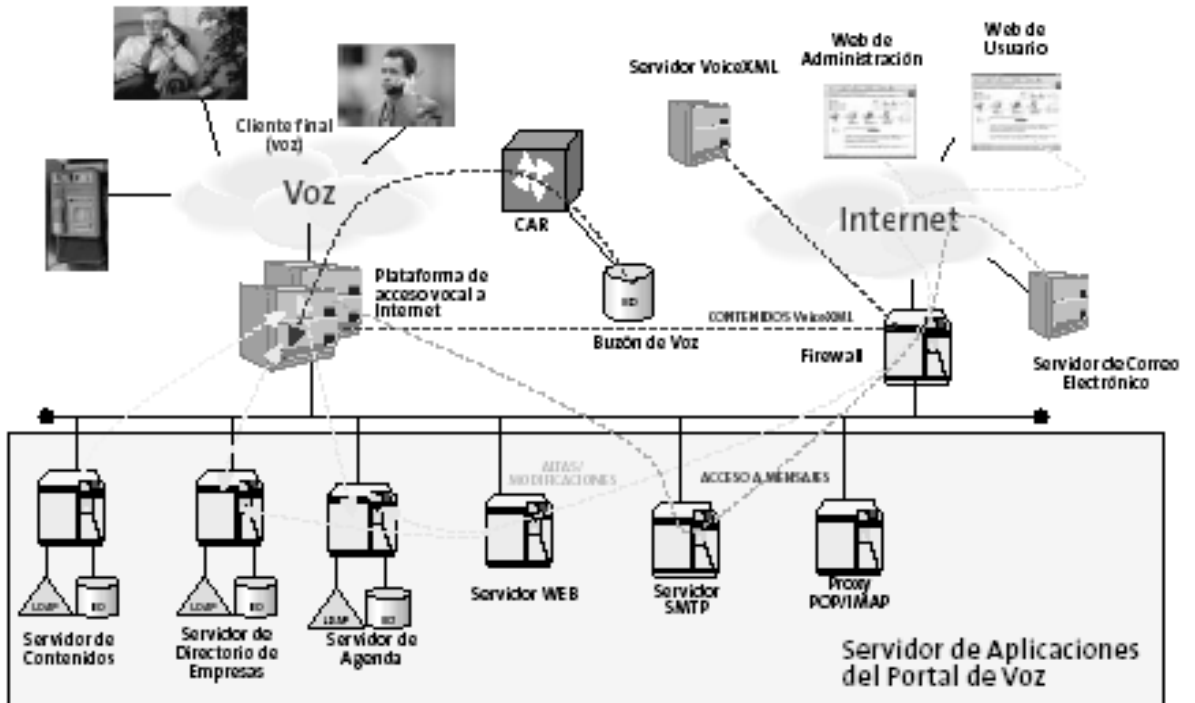
El Servidor de Agenda, con un repositorio de direcciones particular de cada usuario que permite el envío de correo electrónico y la realización de llamadas salientes.

El Servidor de Directorio de Empresas, con un repositorio de información de empresas para el envío de correo electrónico y realización de llamadas saliente.

El Servidor Web, que se comunica con el Servidor de Agenda, para suministrar un Web que permita, a cada usuario, darse de alta en el servicio, configurar su servidor de correo y actualizar su agenda particular.

También se comunica con el Servidor de Directorio de Empresas, para suministrar un Web de Administración que permite actualizar el directorio de empresas por parte del operador del sistema.

Figura 9 Diseño del HW y SW del servidor de aplicaciones



Fuente: Portales de Voz: Internet en el teléfono
http://www.tid.es/presencia/publicaciones/comsid/esp/19/ART_2.PDF (2003)

- El servidor de transcodificación

Permite convertir a formato VoiceXML contenidos de cualquier origen (basados en XML). Su funcionamiento es el siguiente:

- Para una URL dada, se busca si existe un patrón de transcodificación definido; de ser así, se aplica y se genera dinámicamente el VoiceXML.
- Para una URL que corresponda con una página WML, la conversión se realiza de forma automática y sin patrones, aplicando reglas de transformación generales. Se obtiene así una página VoiceXML generada también dinámicamente.

Este servicio proporciona a la plataforma contenidos en VoiceXML, que la plataforma transforma automáticamente a su interfaz vocal para proporcionárselos al usuario.

- El módulo de gestión de la transcodificación

Este módulo permite asociar un conjunto de patrones de transformación aplicables a una URL concreta, para facilitar y automatizar la manera en que se realiza la conversión de cada uno de los formatos. Esta gestión se realiza de una forma gráfica y amigable, lo que facilita la definición de dichos patrones.

- *El módulo de personalización*

Con este módulo se permite la generación dinámica de páginas VoiceXML adaptadas para el usuario, bien por personalización explícita del mismo a través del Web del servidor de aplicaciones, o bien implícitamente por “aprendizaje” del propio módulo de personalización. La reproducción de las páginas VoiceXML se realiza mediante el *Browser Vocal*.

Junto a estos bloques existen otros dos funcionales que sirven de apoyo a la generación y gestión de contenidos:

1. El Entorno de Desarrollo VoiceXML

Proporciona un entorno de diseño de contenidos en VoiceXML, que la plataforma ya es capaz de interpretar y ofrecer al usuario de forma vocal. Para ello, incorpora un editor gráfico de VoiceXML que evita al usuario el conocimiento del lenguaje de marcas asociado, así como la posibilidad de probar las páginas generadas con la herramienta sin necesidad de ser interpretadas por la plataforma.

2. La gestión de contenidos

Es fundamentalmente la alternativa a la transcodificación de contenidos. En lugar de tener un tipo de páginas y poder generar distintos tipos mediante transformaciones, se parte de unos contenidos únicos. Típicamente, con base de datos y mediante plantillas de fácil programación, se generan páginas en cualquier lenguaje: html, wml, VoiceXML.

Con este tipo de plataforma se pueden atender todos los tipos de llamadas de los usuarios que requieran acceso a contenidos, con independencia del origen y del formato en el que se encuentren dichos contenidos.

Algunos casos prácticos de atención son:

- ✓ *Acceso a contenidos html/wml en Internet:*
 - a) El usuario llama al servicio desde un teléfono.
 - b) Es atendido por el Nodo de Servicios Multimedia (encargado de poner las locuciones, hacer el reconocimiento de lenguaje natural e implementar la lógica del servicio como tal) dándole la bienvenida al Portal de Voz y obteniendo su perfil de usuario.
 - c) El usuario solicita un contenido *html* o *wml* que reside en Internet.
 - d) El Nodo de servicios multimedia lanza una petición de recogida de página al *browser* vocal, y este, tras acceder al contenido, lo “cachea” para futuras peticiones y lo transmite al servicio de transcodificación.
 - e) El servicio de transcodificación realiza la conversión del contenido a VoiceXML con la ayuda del gestor de transcodificación.
 - f) El servicio de transcodificación pasa el contenido VoiceXML al módulo de personalización. Este realiza la adecuación de los contenidos al cliente, si es necesario.

g) Finalmente, el módulo de personalización pasa los contenidos al intérprete VoiceXML del browser vocal, que manda las órdenes de puesta de locuciones, recogidas al nodo de servicios en función del contenido VoiceXML final.

✓ *Acceso a contenidos VoiceXML en Internet:*

- a) El usuario llama al servicio.
- b) Es atendido por el nodo de servicios multimedia, que le da la bienvenida y obtiene su perfil.
- c) El usuario solicita un contenido escrito en VoiceXML.
- d) El nodo de servicios solicita al browser vocal el contenido; este accede a Internet, y al reconocerlo en formato nativo VoiceXML, lo envía directamente al módulo de personalización.
- e) El módulo de personalización filtra los contenidos si el perfil del cliente lo requiere, enviándolo al intérprete VoiceXML del browser vocal, que transmite la página al usuario final a través del nodo de servicios.

✓ *Acceso a contenidos no XML en Internet y contenidos propios:*

- a) El usuario accede al servicio a través de una llamada telefónica.
- b) Es atendido por el Nodo de Servicios Multimedia, que le da la bienvenida y obtiene su perfil.
- c) Si el contenido que solicita es un contenido en Internet no escrito en VoiceXML, por ejemplo un correo electrónico, el nodo de servicios lanza una petición de recogida del contenido al servidor de aplicaciones.
- d) El servidor de aplicaciones accede como corresponda al contenido, por ejemplo, por el protocolo POP3, y lo suministra al nodo de servicios.
- e) El Nodo de Servicios realiza los filtrados adecuados de corrección de errores ortográficos típicos y de terminología de Internet, para

posteriormente suministrarlo al usuario mediante conversión texto a voz.

f) Si, por el contrario, el contenido está en formato propio en una base de datos del sistema o en cualquier otro formato, el nodo de servicios realiza el acceso, bien directamente o bien a través del servidor de aplicaciones, para obtener el contenido y proporcionárselo al usuario.

Integración con la plataforma telefónica

El servicio de portales de voz, como se ha estado observando, necesita de una plataforma telefónica, la cual sirve de intermediaría entre el usuario y el servicio.

La nueva tecnología VoIP, mencionada anteriormente, permite que la integración de esta plataforma sea mucho más fácil debido a que permite tener una conexión directa con la red IP. Por otro lado, el poder prestar servicios telefónicos sobre estructuras de red basadas en protocolo IP disminuye considerablemente los costes de la plataforma y del servicio.

Analogía con los portales visuales (navegación convencional y navegación vocal)

Algunos de los términos utilizados en el entorno de Internet, necesarios para proporcionar este servicio, se siguen conservando o se ha diseñado el equivalente como en los casos mostrados en la tabla:

Tabla 1. Internet tradicional contra internet por voz

Internet tradicional	Internet por Voz
Navegador Web	Browser vocal o navegador por voz
Páginas Web	Audiopáginas
Portal Web	Portal de voz
Html	VoiceXML

Sin embargo, se debe saber que cualquier empresa que disponga del personal que conozca sobre Java, *Javascript*, CGI's, *servlets*, LDAP, etc.) puede construir un sistema para obtener el correo por voz utilizando VoiceXML ya que es muy sencillo hacerlo.

La forma de navegar mediante este nuevo servicio, aunque se encuentra bastante restringida, conserva siempre la típica característica de la navegación convencional: los menús.

Forma de uso

La forma de utilizar este servicio es realmente sencilla. Solo se debe llamar al número indicado y en seguida se tendrán las opciones de consulta; en otras palabras, el menú, para lo cual sólo se debe pronunciar la palabra que corresponda al menú deseado. En algunas ocasiones el sistema suele preguntar algún dato o número concreto que necesite para conseguir la respuesta. Luego el portal va delimitando el recorrido del usuario mediante un sistema de preguntas y respuestas. Sin embargo, también pueden surgir inconvenientes, debido a que la información disponible es mucho menor que la que se puede encontrar al utilizar una PC. Además, la tecnología de reconocimiento de voz es muy delicada y no funciona bien si hay excesivo ruido de fondo.

Actualmente algunos de los portales de voz se mantienen con publicidad, la cual es presentada cada vez que se solicita un servicio. La manera de hacerlo consiste en que el usuario escuche quién es el anunciante gracias al cual puede disponer de la información gratuitamente. Sin embargo, los portales de voz aspiran a que su fuente de ingresos sea, en un futuro próximo, el cobro de comisiones por las compras que realicen los usuarios a través de su servicio.

Como se dio a conocer anteriormente, los portales de voz pretenden por el momento resolver necesidades cotidianas. Ofrecen lectura de correo, información del tiempo, cotizaciones, deportes, tráfico y datos similares. Un ejemplo típico podría ser una persona que va viajando en su carro y, con el deseo de encontrar un hotel por la zona para poder descansar, al acceder al portal este determinará automáticamente la procedencia de la llamada y seleccionará para él los lugares más cercanos.

Es importante mencionar que cuando el usuario hace la llamada telefónica, el sistema convierte la palabra en un código binario, el cual es utilizado por la tecnología digital para buscar la petición de la información, que será transmitida al usuario como si al otro lado de la línea hubiera un teleoperador.

“El portal de voz forma parte de la creciente tendencia que desdibuja las diferencias entre computadoras, teléfonos celulares y los distintos híbridos que existen en el mercado”⁴.

Pero debido a la información limitada de la Web que pueden brindar, la mayoría de este tipo de portales han sido creados con menús restringidos para poder encontrar la información con facilidad.

Objetivos de los portales de voz

Los objetivos prioritarios de los portales de voz son:

- La universalización del concepto Internet.
- El incremento de tráfico telefónico.
- El incremento de la masa crítica de usuarios de acceso vocal a Internet por teléfono.
- La proliferación de nuevos negocios asociados: *e-commerce*, juegos interactivos, facturación, publicidad, etc.

Ventajas de los portales de voz

Las ventajas que este tipo de portales promete brindar son:

- La operación por voz permite a los usuarios acceder a información y realizar transacciones en forma natural, ágil y automática
- Elevan la productividad.
- Reducen los costos operativos.
- Mejoran la calidad de servicio al cliente, incrementando su satisfacción.
- Generan novedosos servicios diferenciales.
- Reutilizan las inversiones ya realizadas en infraestructura y contenidos Web.
- Funcionan en forma automática 24 horas por día.
- Son accesibles a través de múltiples canales (fax, e-mail, sms, teléfono) y estándares (MMS, WML, 2.5G, 3G).

En la actualidad aún no se pueden apreciar todas las ventajas de estos portales, debido a que las aplicaciones son escasas. Sin embargo, en cuanto la seguridad en las transacciones orales esté plenamente garantizada, estos portales estarán dentro del mercado del comercio electrónico, con lo que se lograrán varias ventajas, como prescindir de pantallas, y ya no tener que teclear, ya que sólo con hablar se podrá realizar la transacción. Por otra parte los portales de voz constituyen una gran ventaja para las personas que no pueden leer o escribir, por lo que en los países menos desarrollados este servicio sería muy útil.

Una de las principales ventajas de los portales de voz, hoy en día, es que tratan por todos los medios de descubrir qué cosas son las que más necesitan las personas.

Las ventajas que todo usuario puede observar son:

- Se puede acceder desde cualquier teléfono
- No hay barreras de conocimiento
- No se necesita aprender complejos comandos
- Se puede reconocer cualquier tipo de voz

Es importante saber que existe una ventaja significativa en cuanto a los sistemas de verificación de voz, y es que estos permiten identificar a las personas con un factor de seguridad diez veces superior al de la huella digital.

Desventajas de los portales de voz

La mayoría de servicios proporcionados siempre tienen desventajas y este servicio no es la excepción; entre las principales están:

- Limitación de la cantidad de información que se puede distribuir
- El ruido de fondo puede llegar a distorsionar la información que se desea comunicar al usuario

Viabilidad

En los países desarrollados, el número de compañías que se preparan para tener presencia en el mercado de los servicios de voz de Internet crece día a día. Todas se apoyan en el formato VoiceXML, sobre el cual se ampliará en el siguiente capítulo.

Los países donde predomina este tipo de servicio son Estados Unidos y España. TellMe.com e Ydilo.com son una muestra de este tipo de portales. Según estudios hechos por la consultora *Gartner Group* se considera que actualmente un 30% de los proyectos de este tipo están en funcionamiento y el crecimiento mayor se ha dado en los portales de voz que brindan servicio de noticias, pronósticos del tiempo, cotizaciones de la bolsa, correo de voz. Sin embargo, el mercado de los portales de voz espera extenderse al comercio electrónico, para lo cual se está trabajando en alcanzar los niveles de seguridad requeridos.

“*The Kelsey Group* prevé que en el 2005 habrá en el mundo 18 millones de usuarios de algún tipo de portal de voz”⁵.

Según estadísticas que se han hecho, el número de usuarios de Internet crece a un ritmo superior al 100% anual, y la telefonía móvil lo hace

a un ritmo entre el 60-80%. Por tal motivo, no existe duda alguna de que al converger estos dos sectores las previsiones son muy optimistas.

En los países desarrollados la mayor parte de empresas de telecomunicaciones, fabricantes de terminales, empresas proveedoras de contenidos y las grandes marcas han observado estos porcentajes de crecimiento, por lo que ven en esta forma de acceder a Internet una gran oportunidad para los negocios, por lo que han comenzado a hacer alianzas e inversiones conjuntas. La aparición de estos nuevos actores está siendo decisiva en la mejora y el desarrollo de los servicios de los portales de voz.

En Guatemala aún existe poco conocimiento sobre los portales de voz. Las empresas telefónicas que conocen al respecto los ven como una alternativa lejana, y quienes ya lo incorporan lo hacen únicamente con los servicios básicos y a corta escala y, obviamente, está disponible solo para sus clientes, por lo que se puede considerar que en nuestro país este tipo de servicio se encuentra en etapas de estudio.

3. EL LENGUAJE ESTÁNDAR PARA EL ACCESO ORAL A INTERNET

3.1 Historia

El lenguaje HTML por sencillo que parezca es uno de los mejores inventos. “Es el más exitoso sistema de presentación de documentos de la historia”⁶. Gracias a este lenguaje se ha podido publicar y acceder a un sin número de información de la Web.

De todos es sabido que Internet ha tenido un gran crecimiento por lo que los intereses comerciales y la necesidad de poder presentar páginas vistosas en la Web han ocasionado que el HTML evolucione rápidamente. Sin embargo, este sigue siendo igual de rígido e inflexible como en sus inicios.

Por tal motivo, la W3C ha desarrollado el XML que aprovecha las ventajas del HTML, pero permite realizar varias cosas más. Este en realidad es un metalenguaje. Los navegadores todavía no soportan en toda su potencia el XML y sus tecnologías asociadas, pero se espera que en el futuro esto cambie.

VoiceXML es un lenguaje que parte de la familia XML (Extensible Markup Language). Nació en los laboratorios Bell de AT&T, con el proyecto PhoneWeb. Sin embargo, debido a la división de esta compañía en AT&T, por un lado, y Lucent Technologies, por otro, el proyecto continuó en Lucent, esta vez con el nombre de TelePortal.

En 1998 la empresa Motorola creó el VoxML (Voice Markup Language) lo que constituyó otro aporte para VoiceXML, y sirvió como una base para él. Años más tarde, en unión con AT&T, Lucent, IBM y W3C (World Wide Web Consortium) crearon el VoiceXML Forum, sitio que sigue la evolución detallada de la tecnología de reconocimiento de voz en Internet. Este lenguaje se ha convertido en el estándar para la creación de navegadores y servicios online a los que se puede acceder desde un teléfono convencional.

VoiceXML unifica la forma de representar los contenidos Web siguiendo las pautas de XML. De esta forma es posible crear documentos de forma automatizada, que definen cómo y qué información presentar al usuario vía voz.

Desde que el W3C aprobó la especificación 1.0 del XML, en febrero de 1998, se ha generado una gran cantidad de información aplicaciones y software alrededor de este nuevo estándar.

3.2 Definición

VoiceXML (Voice Extensible Markup Language) es un lenguaje extensible de marca de voz, el cual proporciona una interfase de programación de alto nivel de recursos de voz y telefonía para los proveedores de servicios y desarrolladores de aplicaciones. El VoiceXML sigue todas las reglas sintácticas de XML, con las semánticas, que son las que apoyan la creación de aplicaciones de voz interactivas. Todas las estandarizaciones establecidas para este lenguaje permitirán el acceso por voz a bases de datos, informaciones y servicios de sitios Web e Intranet de las empresas, así como crear nuevas aplicaciones y dispositivos. Con este lenguaje se espera poder extender Internet a los teléfonos y a otros dispositivos utilizando únicamente la voz.

Este estándar es gestionado por el VoiceXML Forum. La base de este lenguaje son aplicaciones IVR y ASR, lenguaje diseñado para hacer disponible la información de Internet a través de un teléfono.

VoiceXML o VXML, al igual que HTML, es un lenguaje markup renderizado por un navegador de voz, así como HTML es renderizado por su navegador Web para formatear el contenido y formularios de entrada de datos.

Las aplicaciones desarrolladas mediante VXML usan habla sintetizada o archivos de audio pre grabados. Su software puede recibir la entrada del usuario usando el habla o por tonos del teclado de telefónico.

VXML se integra con motores de reconocimiento de voz y como se ha venido mencionando, soporta el acceso telefónico vía Web, la navegación en Internet por medio de teclas y el reconocimiento de voz.

Con VoiceXML, cualquier sitio Web puede añadir con facilidad un acceso telefónico a sus contenidos y servicios

Casi todos los portales de voz se basan en este lenguaje. Debido a su facilidad y parecido al HTML, los proveedores de contenidos pueden abrir sus servicios Web a los clientes de teléfono sin mayores dificultades. Además, los programadores pueden construir servicios automáticos de voz utilizando el mismo tipo de software que se usa para crear Webs visuales.

3.3 World Wide Web Consortium

El World Wide Web Consortium (W3C) ha creado una serie de grupos de trabajo con el fin de cumplir la tarea de extensión de VoicexML. Cada grupo se encarga de diferentes aspectos que permiten, al final, consolidar una serie de estándares para dicho lenguaje.

Estos grupos de trabajo son los siguientes:

1. *Speech Grammars Group*. Es el encargado de definir la sintaxis de las gramáticas que se usarán en el reconocimiento de voz, trabajo que está bastante avanzado y del cual se cuenta con un borrador de la especificación. También este grupo es el encargado de trabajar en la definición de modelos de lenguaje estocásticos, que permitirá la definición de grandes vocabularios e incluso de vocabularios abiertos.

2. *Voice Dialogs Group* Este grupo tiene como tarea la realización de extensiones y modificaciones al actual VoiceXML, para mejorar el lenguaje de definición de la interacción vocal con los servicios.

3. *Speech Synthesis Group* La misión de este grupo es definir un lenguaje de marcas, basado en XML para la voz sintética, mediante el cual se puedan manipular, de forma sencilla e intuitiva, los parámetros básicos del sintetizador de voz, como pueden ser la entonación, el volumen, la velocidad, el tono de la voz sintetizada, etc.

4. *Pronunciation Lexicon Group*. Este grupo está encargado de definir un formato para la representación de palabras o expresiones (abreviaturas, expresiones extranjeras) con cuya pronunciación se construye la voz sintética.

5. *Call Control Group*. La misión de este grupo es especificar los requisitos de acceso telefónico para, posteriormente, definir el formato de representación del manejo de recursos de telefonía en plataformas VoiceXML, utilizando recursos periféricos de la red y sin entrar en el núcleo de la misma.

6. *Multimodal Systems Group*. Este grupo se encarga de definir los requisitos y procedimientos para la ampliación de las interfaces vocales definidas con VoiceXML a nuevos medios de interacción, tales como teclados y pantallas gráficas. La aportación de este grupo resultará, por tanto, crucial para la estandarización de nuevos lenguajes de presentación e interacción multimedia.

7. *Reusable Dialog Components Group*. Este grupo está especificando los requisitos para la definición e invocación de diálogos reusables, que puedan ser implementados por expertos en las plataformas de tecnología del habla, e incorporados como librerías, pudiendo ser, por tanto “invocados” por los creadores de los servicios.

Como puede verse, todos estos grupos están desarrollando actividades para la extensión del lenguaje VoiceXML, incluso en áreas no cubiertas por la especificación actual y no relacionadas muy directamente con las posibilidades multimedia de los servicios, como es el caso de aquellas actividades derivadas de los nuevos modelos de red (control de llamada) y de las interfaces multimedia.

3.4 Sintaxis

Este lenguaje al igual que su 'padre' XML, está basado en “*tags*” (etiquetas), que encierran instrucciones llamadas “*item*” (elemento), como se podrá observar en los ejemplos descritos posteriormente. Un *tag* puede tener atributos; cada atributo consta de un nombre y un valor, separados por el signo igual (=); además, el valor debe estar encerrado entre comillas. En la siguiente tabla se puede observar la estructura de VXML.

Tabla II. Estructura Básica de Voice XML

```

< nombre_elemento
nombre_atributo=
"valor_atributo">
.....
..... Ítems
< /nombre_elemento>
    
```

Entre las etiquetas comunes que se utilizan se encuentran <vxml>, <form>, <var>, <menu>, etc.

Los documentos escritos en VoiceXML deben llevar como primera instrucción la etiqueta que indique la versión XML; por ejemplo: <?xml versión="1.0"?> Luego debe ir la etiqueta <vxml> y el atributo de la versión correspondiente, de esta forma: <vxml version= "2.0">, y al final de las instrucciones debe ir la etiqueta: </vxml>

Existen diversos tipos de instrucciones que se manejan en VoiceXML, por lo cual se describirán los más sobresalientes con sus respectivos ejemplos:

- Diálogos VoiceXML:

Los diálogos sirven de interface con el usuario. Se definen dos tipos: form y menú

- form: sirve para presentar información al usuario u obtener información de él.
- menu: se puede considerar como una *form* especializada que "fuerza" al usuario a seleccionar una opción y extenderse por ésta.

- Elemento de una form:

Llamados “form Ítems”, son todos aquellos elementos que pueden estar contenidos en una form, en otras palabras, bajo la etiqueta <form> se clasifican:

- field items: sirven para recoger información del usuario y complementar los campos de variables, ejemplo: un prompt que indique al usuario que seleccionar;
 - control items: comprenden tareas no basadas en reconocimiento
- Field items:
 - <field>: recoge información a partir del usuario a través de reconocimiento de voz o de DTMF
 - <record>: registra un clip de audio del usuario
 - <transfer>: transfiere al usuario a otro número telefónico
 - <object>: invoca un objeto de una plataforma específica que puede recoger la información del usuario retornando el resultado con un objeto ECMAScript
 - <subdialog>: desarrolla una llamada a otro diálogo o documento (similar a una función call) retornando el resultado como un objeto ECMAScript

- Elementos de control:

Denominados “Control items”.

- <block>: engloba una secuencia de sentencia para "prompting" y cómputo
 - <initial>: controla las interacciones "mixed-initiative" en una form
- Otros elementos:

- `<prompt>`: elemento que controla la salida de voz sintetizada y audio pregrabado, como forma de "presentación" del sistema. Puede ser omitido explícitamente solo cuando:
 - no se requiere especificar alguno de sus atributos, y
 - consiste totalmente de PCDATA (no contiene marcas de voz), o de un elemento `<audio>`. marcas de voz: "marcas" que pueden contener los prompts, para pausas, énfasis, etc.

Ejemplo: utilizando `<form>`, `<block>` y `<prompt>`

```
<?xml version= "1.0">
<vxml application= "ejemplo.vxml" version "1.0">
  <form id="someName">
    <block>
      <prompt>
        Buenos Días
      </prompt>
    </block>
  </form>
</vxml>
```


Ejemplo del uso de prompt desde fichero “wav” pregrabado

```
<?xml version "1.0"?>
<vxml application="ejemplo de prompt.vxml" version="1.0">
  <form id="someName">
    <block>
      <prompt>
        <audio src="www.greeting1.net">
      </prompt>
    </block>
  </form>
</vxml>
```

- <audio>: etiqueta de “presentación” perteneciente al elemento prompt, típicamente se especifica a través de un URI. También puede estar especificada en una variable de audio pregrabada.

- Aplicaciones Multi-Form

Se necesita para desarrollar aplicaciones IVR o ASR, pues requieren:

- implementar control de diálogos
- desarrollar tareas de reconocimiento
- manipular eventos, etc.

Ejemplo:

```
<form id= "Obtener_numero_telefónico">
  <field name="Número_telefónico">
    <prompt>
      ¿Cuál es su número de teléfono?
    </prompt>
    <grammar src=" ../grammars/telefono.gram"
      type="application/x-jsgf"/>
    <help>
      Por favor diga los siete dígitos de su número telefónico
    </help>
  </field>
</form>
```

- Gramática

Representada por la etiqueta <grammar>, se emplea para proveer una “gramática de conversación o del habla” de manera tal que:

- especifique un grupo de palabras que el usuario puede decir para desarrollar una acción o suministrar información
- provea una secuencia de valores (caso de grammar field) o grupo de parejas valores-atributos (caso de form grammar) para describir la información o acción

-Grammar inline:

```
<grammar type="mime-type">
  inline speech grammar
</grammar>
```

-Grammar externo:

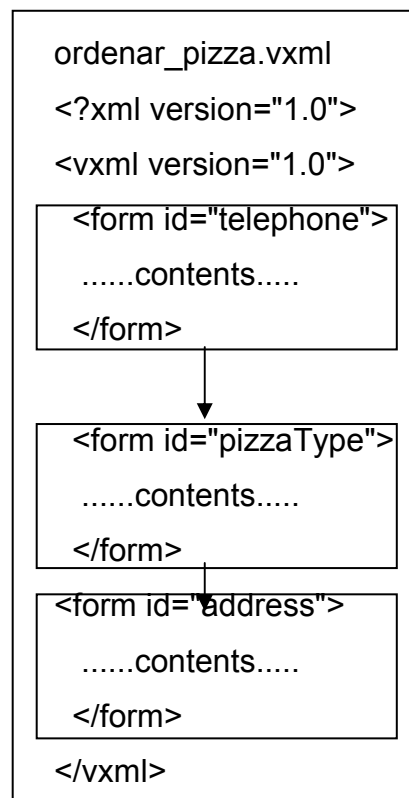
```
<grammar src= "URI" type="mime-type"/>
```

- Tránsito entre forms

Normalmente las aplicaciones necesitan múltiples forms dentro de un documento VoiceXML, por lo que se requiere de un "mecanismo" de transición entre las forms. Típicamente se usa la etiqueta <goto> para transición entre forms.

Ejemplo: (Aplicación VoiceXML con tránsito entre forms para "ordenar una pizza")

Tabla III. Aplicación VoiceXML con transito entre forms



- Posibles tipos de tránsitos

Con el elemento <goto> se pueden desarrollar diferentes tipos de "tránsitos entre forms":

- Tránsito a otra form en un diálogo (en la form actual)
- Tránsito a otra form (diálogo) en el documento actual

- Tránsito a otro documento

Tránsito a otra form en un diálogo:

```
<goto next item="some_form_items_var_name">
```

Tránsito a otra form (diálogo) en el documento actual

```
<goto next="#some_form_id">
```

Tránsito a otro documento

```
<goto next="http:www.some_url.com/some_doc.vxml"/>
```

- Uso del atributo "expr":

Es una manera de resolver, dinámicamente, los URI y los fragmentos de URI, utilizando el atributo "expr" y una expresión ECMAScript:

```
<goto expr="../some_app/+filenamevar + .vxml"/>
```

- Sentencias condicionales:

Los elementos utilizados para las sentencias condicionales son:

- <if>
- <else>
- <elseif>

cada elemento debe utilizar un atributo "cond" que especifique una condición.

Ejemplo:

```
<if cond="total > 1000">
  <prompt>
    this is too much to spend
  </prompt>
</if>
```

3.5 Ejemplos vistos desde la perspectiva de ejecución de un *browser* vocal

Se ha podido observar la estructura básica del lenguaje VoiceXML, por lo que en seguida se mostrará un ejemplo del funcionamiento de una aplicación codificada en este lenguaje y ejecutada sobre un *browser* vocal.

En primer lugar para poder acceder a ella el usuario que llama ha marcado un número telefónico convencional, al que se le ha asignado esta página VoiceXML, de manera que, cuando se recibe una llamada en este número, el *browser* va a buscar la URL en donde reside dicha página.

Con esta página vocal o **audiopágina** podría tener lugar un diálogo como el que se describe en seguida:

Browser vocal: *Bienvenido al servicio de Pizza Mía.
Seleccione el tamaño de su pizza:*

Pequeña, mediana o grande.

Usuario: *Grande, por favor.*

Browser vocal: *¿Qué especialidad desea?: Cuatro Estaciones,
Marinera, Barbacoa o Especial.*

Usuario: *Barbacoa.*

El *browser* vocal captura la audiopágina y toma secuencialmente los textos contenidos entre los delimitadores **<prompt>** y **</prompt>**, y los pronuncia utilizando conversión texto a voz. De la misma manera, toma las palabras comprendidas entre los indicadores **<grammar>** y **</grammar>** y construye dinámicamente un vocabulario que es entregado a un reconocedor de voz con las palabras de selección.

Ocasionalmente, este vocabulario será completado automáticamente por el propio *browser* con un conjunto de palabras de control de navegación (por ejemplo, anular, anterior, etc.), que permiten la navegación por las audiopáginas que han antecedido a esta en la interacción del usuario o el borrado del formulario vocal. Una vez que se han obtenido los datos requeridos en esta audiopágina, el *browser* selecciona una nueva URL. Es importante observar el paralelismo entre la aplicación audio y la aplicación visual HTML, accesible de modo convencional.

El lenguaje VoiceXML ha sido diseñado para crear diálogos vocales entre personas y máquinas; en dichos diálogos se utilizan básicamente técnicas de conversión texto a voz, reproducción de voz (señales de audio en general) codificada, grabación y codificación de voz, reconocimiento de voz y detección de teclas DTMF; dado que el acceso a estos diálogos se efectúa a través de la red telefónica, es necesario también usar técnicas de telefonía y de control de llamadas telefónicas. El objetivo último y primordial del lenguaje es acercar al mundo de los servicios de voz, así como la facilidad y la potencia para el manejo de contenidos de los servicios en el mundo Web.

El ejemplo que se mostrará inmediatamente pretende dar a conocer que en una simple llamada se pueden realizar varias tareas, y todo esto debido a la integración de los servicios de voz.

Usuario: Añadir a mi agenda cena el sábado 7 a las 8 p.m. con el Sr.
Escobar

Usuario: Necesito un restaurante italiano en la Capital de Guatemala

Sistema: Se recomiendan 2 restaurantes, desea escuchar sus características?

Usuario: Sí

Sistema: El restaurante La Piazza brinda...

Usuario: Chequee si hay disponibilidad para el próximo sábado a las 8 p.m. en el restaurante La Piazza

Sistema: Hay disponibilidad, se hace la reserva?

Usuario: Sí

Usuario: Abrir mis contactos y llamar al Sr. Escobar

Usuario: Deseo viajar de Esquipulas a Guatemala el próximo sábado

Sistema: Hay disponibilidad en el bus de Rutas Orientales de las 10:00 a.m....

Usuario: Muy bien, deseo reservar un boleto

Sistema: Se necesita su número de tarjeta de crédito y....

Usuario: Sí, mi numero de tarjeta Visa es 654.....

(el sistema verifica los datos de la tarjeta..)

Sistema: Su boleto está reservado.

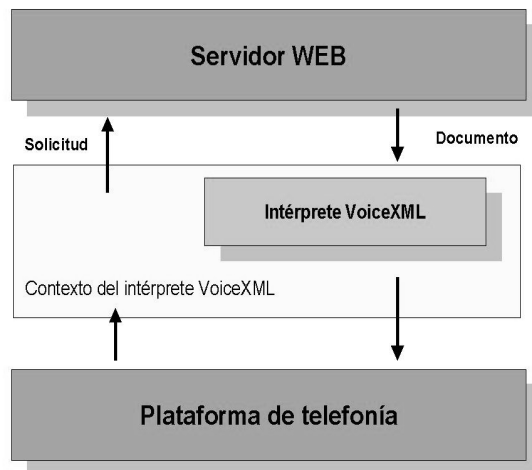
Como se pudo observar en el ejemplo, se realizaron 4 tareas con una sola llamada, gracias a que la integración de estos servicios permite tener acceso a la agenda, a la lista de contactos y a la localización de boletos para viajar y de reservación para un restaurante determinado.

3.6 Modelo arquitectónico del VXML y recursos necesarios para interactuar con el mismo

Como se sabe, la lógica del servicio conserva uno de los principios elementales del desarrollo en Internet: “La arquitectura Cliente-Servidor” (donde los CGIs proporcionan la información y las páginas cliente html se encargan de la presentación). Esto se puede observar en el control del flujo de la llamada el cual es mantenido por las paginas proporcionadas por el servidor de documentos o servidor Web, mientras que la interacción con el usuario es controlada por el browser. Las aplicaciones VoiceXML están constituidas por un conjunto de audiopáginas escritas en dicho lenguaje.

Como se puede observar en la figura #11, la arquitectura básica incluye: la plataforma telefónica, el contexto del VoiceXML con el intérprete VoiceXML embebido y el servidor Web.

Figura 10 Arquitectura básica de las aplicaciones escritas en VoiceXML



Fuente: Tecnología de voz

<http://www.uap.edu.pe/Fac/16/syllabus/16117/Tecnologia%20de%20voz%20-%20semana%206.PDF> (2003)

VoiceXML describe los recursos que una plataforma necesita para interactuar con el usuario

- Voz sintetizada (conversión texto a voz)
- Reproducción de audio
- Reconocimiento de voz
- Reconocimiento de tonos de marcado DTMF
- Grabación de voz
- Funcionalidad de telefonía como transferencia de llamadas o desconexiones

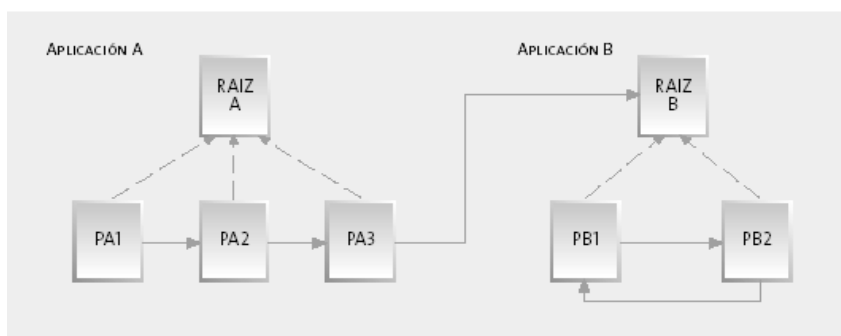
3.7 Contexto de ejecución de aplicaciones VoiceXML

VoiceXML permite que las audiopáginas estén relacionadas entre sí, de manera que los usuarios pueden navegar por ellas e incluso saltar de una aplicación a otra. Cuando se están creando aplicaciones VXML se puede crear una audiopágina-raíz, como se muestra en la figura 12. En esta audiopágina se pueden definir variables, funciones o cualquier otra cosa que, dentro de la aplicación, trascienda la ejecución de una audiopágina y deba ser accesible a otras audiopáginas de la misma aplicación. De manera más general se puede considerar que, al momento de estar interpretando una audiopágina VXML, realmente el browser y su contexto tienen activas dos: la actual y la raíz.

Como se refleja también en la figura 12, cuando se pasa de la aplicación "A" a la aplicación "B", la audiopágina raíz de "A" es eliminada del contexto y la de "B" es inmediatamente cargada e interpretada.

Cada una de las audiopáginas que componen una aplicación está compuesta por un conjunto de diálogos. Cada uno de ellos, junto con la audiopágina raíz, contiene toda la información necesaria para saber cuál es el siguiente diálogo (en la misma audiopágina, en distinta audiopágina de la misma aplicación o en otra aplicación distinta) al que evolucionar, en función de las distintas posibilidades que puede ejercer el usuario de la aplicación.

Estas posibilidades de utilización (ya sea mediante reconocimiento de voz o por detección DTMF) se conocen con el nombre genérico de gramáticas. Estas gramáticas se componen dinámicamente cada vez que se interpreta una página y en esta construcción se tienen en cuenta las palabras (en el caso del reconocimiento de voz) que estén definidas en la audiopágina-raíz.

Figura 11 Contexto de ejecución de aplicaciones VoiceXML

Fuente: Acceso Vocal a contenidos Internet

<http://www.tid.es/presencia/publicaciones/comsid/esp/20/4XX.PDF> (2003)

3.8 Manejo de entradas y salidas de datos

Las entradas por parte de los usuarios son reguladas mediante las gramáticas, ya que estas constituyen el mecanismo propuesto por VoiceXML. Así mismo, para poder responder a determinadas situaciones se cuenta con la definición de eventos, los cuales son generados por la plataforma de implementación y actúan según las situaciones que puedan suceder, como por ejemplo: si el usuario no dice nada o si el reconocedor de voz no entiende lo que el usuario dijo, si el usuario cuelga, etc.

Existe un conjunto de eventos que VoiceXML debe definir para que el browser proporcione tratamientos generales, con lo cual el proveedor de aplicaciones quedaría libre de invertir tiempo definiendo este tipo de eventos. Sin embargo, también el proveedor tendrá acceso a manejar dichos eventos según lo requiera. Por lo que cuando se produce un cierto evento, el browser verificará si el tratamiento del mismo está determinado dentro del diálogo en el que se haya producido y, si no lo está, irá escalando en la sucesión jerárquica de la aplicación. Primero, a la audiopágina en que se encuentra incluido el diálogo y después a la audiopágina-raíz.

Cuando se encuentre una acción definida para dicho evento en la cadena anterior, se ejecutará, y si no se encuentra, será ejecutada la acción por defecto que debe soportar la plataforma de implementación.

VoiceXML aporta componentes para emitir locuciones pregrabadas. Pese a ello el medio más apropiado para generar información hacia el usuario es la conversión texto a voz. Para darle mayor realce a esta característica del lenguaje se han incluido marcas específicas para inspeccionar algunos aspectos en la lectura de los textos. De esta forma se pueden controlar el volumen, la velocidad de lectura, el tono grave o agudo de la voz e incluso la forma en que el intérprete debe producir determinadas cadenas de caracteres.

Por ejemplo, no es lo mismo leer una secuencia de números, si se refieren a una cantidad, que si se está proporcionando un número de teléfono; por ello, se han determinado marcas específicas que, ante una misma secuencia de caracteres, pueden provocar una salida de audio distinta. Al mismo tiempo, se provee también un mecanismo para que las facilidades específicas de una determinada plataforma de implementación puedan ser usadas por los proveedores de aplicaciones. Evidentemente, el uso de estas extensiones debe limitarse, en la medida de lo posible, con el objetivo de mantener la portabilidad de las mismas cuando se ejecutan en browsers de distintos suministradores.

Uno de los puntos más débiles de este lenguaje es el proceso que se propone para realizar funciones telefónicas, ya que lo único que se maneja es el hecho de que el usuario se conecte mediante una llamada telefónica, pero no se ha previsto ningún procedimiento para que el servidor de documentos (que es donde se encuentran las audiopáginas alojadas) pueda lanzar una llamada saliente de forma autónoma.

Tampoco es posible generar una llamada saliente sobre la que se pueda ejecutar una aplicación VoiceXML, antes de conectar con el usuario que llama: este esquema es habitual en los servicios telefónicos de cobro revertido, donde el usuario llamado tiene la facultad de aceptar o rechazar la llamada entrante.

Los ejemplos que a continuación se presentan involucran mayor contenido en cuanto a código y se comportan de manera diferente dependiendo de la respuesta del usuario, y ponen en evidencia lo descrito en cuanto a entradas y salidas se refiere. En el segundo ejemplo se podrá observar cierto tipo de gramática utilizada para el tratamiento de las entradas por parte del usuario.

Ejemplo # 1: cuestionario sobre gustos de animales

```
<vxml>
  <form id="cuestionario_animal">
    <field name="animal_favorito">
      <prompt>
        <audio>¿Cuáles prefieres mas, los perros o los gatos?</audio>
      </prompt>
      <grammar>
        <![CDATA[
          [
            [perros chuchos cachorros perritos chuchitos
              cachorritos] {<option "perros">}
            [gatos mininos gatitos] {<option "gatos">}
          ]
        ]]>
      </grammar>
    </field>
    <!-- si el usuario da una respuesta válida, se ejecuta el bloque
    "filled" -->
```

```
<!-- Tellme solo acepta un campo por diálogo, por eso no hay
ambigüedad acerca del campo al que hace referencia -->
<filled>
  <result name="perros">
    <goto next="#hechos_de_perros_populares"/>
  </result>
  <result name="gatos">
    <!-- llaves alrededor del nombre de una variable nos da su
    valor -->
    <goto
    next="evaluacion_psicologica.cgi?enfermedad={animal_favorito}"/
    >
  </result>
</filled>
<!-- si la respuesta del usuario no concuerda con la gramática
(grammar), el bloque "nomatch" es ejecutado -->
<nomatch>
  <audio>Lo siento, no entendí lo que dijo.</audio>
  <reprompt/>
</nomatch>
<!-- si no hay respuesta en unos segundos, el bloque "noinput" es
ejecutado -->
<noinput>
  <audio>Lo siento, no le escuche</audio>
  <reprompt/>
</noinput>
</form>
<!-- formularios adicionales pueden ir aqui -->
</vxml>
```

Ejemplo # 2: solicita al usuario un número de 1 a 10. Este ejemplo usa una gramática con deletreos fonéticos de los números para determinar qué dijo el usuario. Luego esta le dice al usuario qué dijo (este ejemplo está en idioma inglés).

Código fuente:

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE vxml PUBLIC "-//Nuance/DTD VoiceXML 1.0//EN"
'http://voicexml.nuance.com/dtd/nuancevoicexml-1-2.dtd' >

<vxml version="1.0">
  <form id="MenuPrincipal">
    <field name="EnteroHablado">
      <!-- Please say a number from 1 to 10. -->
      \!voice="Spanish-Female1" Por favor diga un número de 1 a 10.
      <grammar scope="dialog">
        <![CDATA[[
          [oonoh]      { <SpokenInteger "1"> }
          [dohs]      { <SpokenInteger "2"> }
          [trace]     { <SpokenInteger "3"> }
          [quawtrow]  { <SpokenInteger "4"> }
          [sinko]     { <SpokenInteger "5"> }
          [says]      { <SpokenInteger "6"> }
          [seeaytay]  { <SpokenInteger "7"> }
          [ohchoh]    { <SpokenInteger "8"> }
          [nuwayvay]  { <SpokenInteger "9"> }
          [deeyays]   { <SpokenInteger "10"> }
        ]]]>
      </grammar>

    <noinput>
```

```
<!-- I didn't hear anything. -->
\\voice="Spanish-Female1" No escuch&#233; nada.
<reprompt/>
</noinput>
<nomatch>
<!-- I didn't understand. -->
\\voice="Spanish-Female1" No entend&#237;.
<reprompt/>
</nomatch>
</field>
<filled namelist="EnteroHablado">
  <!-- You said EnteroHablado. Thank you for your response. -->
  \\voice="Spanish-Female1" Usted dijo <value
  expr="EnteroHablado"/>. Gracias por su respuesta.
  <goto next="#MenuPrincipal"/>
</filled>
</form>
</vxml>
```

3.9 Transcodificación de HTML o WML a WXML

Como se ha venido mencionando, el VoiceXML es un lenguaje relativamente nuevo, por lo que aún no se ha dado una explotación comercial de browsers para el mismo y casi todos los existentes están disponibles en el mercado norteamericano por lo que las tecnologías del habla incorporadas están en inglés. No obstante, en la medida en que existan más browsers, el número de aplicaciones hechas en VoiceXML aumentará. Este lenguaje es bastante sencillo y cuenta con la ventaja de estar cercano a HTML o WML, lenguajes de uso común entre los proveedores de contenidos en Internet.

Consecuentemente, es de esperar que, en un futuro no muy lejano, los proveedores de contenidos se deban plantear no solo la generación de su Web, sino la de su audio Web, como está sucediendo hoy en día con los portales y las aplicaciones WAP.

Actualmente hay diversidad de aplicaciones que se generan utilizan HTML (y en algunos casos WML), por lo que es necesario planear la generación de aplicaciones VoiceXML a partir de estos formatos, a través de técnicas de transcodificación dinámica. Lo que se quiere lograr al utilizar estas técnicas es que a partir de una página HTML o WML (estáticas) se produzca una audiopágina VXML que tenga los mismos contenidos y que por lo tanto se adapte al entorno auditivo. Es indispensable tomar en cuenta que esta labor conlleva cierta complejidad, porque una gran parte de la información que observamos en la pantalla de nuestra PC al conectarnos a Internet es visual y orientada a ser manipulada utilizando un teclado y/o un ratón. Así mismo se debe tener en cuenta que cuando el diseño de la página cambia la transcodificación de cada pantalla requeriría una actualización, lo que involucra mas trabajo.

Cuando se piensa en transcodificar paginas WML se observa que el trabajo será un poco más sencillo, pues el hecho de que se tengan menos capacidades de presentación en las pantallas permite que las imágenes incorporada muestren una información mas concisa y dirigida. Sin embargo, es necesario mencionar que por ello las aplicaciones WAP utilizan el conocimiento que el usuario tiene sobre los servicios solicitados.

En la figura 12 se pueden apreciar ejemplos de pantallas WEB y WAP con la misma información. Para el caso de Web se interactúa con la información mediante un código de colores (lo cual no se puede utilizar en un entorno auditivo), y para WAP se obtiene información de un código sobrentendido por el usuario (debido al conocimiento del servicio), el cual puede interpretar sin mayor problema el significado de las claves "C:" y "R:".

Figura 12. Transcodificación dinámica de contenidos WEB y WAP



Fuente: Acceso Vocal a contenidos Internet

<http://www.tid.es/presencia/publicaciones/comsid/esp/20/4XX.PDF> (2003)

En una situación más general sería necesario entrar en el propio contenido semántico de las páginas debido a que las instrucciones de navegación tipo “pulse aquí” no son manejables en un entorno auditivo. Por todo lo expuesto, las técnicas de transcodificación dinámica se deben considerar como un paso intermedio antes de la generación de contenidos formateados particularmente para VoiceXML. Así mismo, los programas desarrollados para este efecto se deben someter a constantes revisiones y mejoras y principalmente dependerán siempre de posibles rediseños de las páginas visuales.

3.10 WAP contra VoiceXML

Como se ha observado, las aplicaciones WAP últimamente se han tornado populares, a tal grado que los diversos proveedores de contenidos en Internet se interesen en crear páginas WML. Y como se mencionaba se espera que las audiopáginas en pocos años cuenten con el grado de importancia y popularidad que las WML tienen actualmente. Por ello, en la siguiente tabla se presenta el contraste entre el uso de WAP y VoiceXML.

Tabla IV. WAP contra aplicación en VoiceXML

WAP	VoiceXML
Requiere teléfonos con WAP habilitado	Se puede usar con cualquier teléfono
El ingreso de caracteres con teclado numérico incómodo	Ingreso mediante voz o teclado numérico
Funciona bien en ambientes ruidosos	Difícil de usar en ambientes ruidosos
Necesita desarrollar versiones de su software para una variedad de gateways WAP	Solo necesita desarrollar una versión de su software
Funciona bien para desplegar listas largas de información	Funciona mal para dar al usuario listas largas de información
El usuario puede ingresar información arbitraria	El usuario solo puede decir frases predefinidas (o puede grabar su habla en bruto, pero esto tiene menos utilidad para programas de computadoras)

3.11 Ventajas y desventajas de VoiceXML

Este lenguaje presenta varias ventajas al momento de desarrollar sistemas de telefonía con interacción por voz. Por ser un lenguaje procedente de XML permite:

- Minimizar el número de interacciones cliente/servidor al poder definir varias en un mismo documento
- Crear servicios de diferente complejidad de forma más sencilla y estructurada
- Generar dinámicamente de documentos que incluyan información almacenada en bases de datos corporativas o contenidos de Internet
- Desde un punto de vista del desarrollador de servicios, VoiceXML oculta la mayoría de las dificultades que presentan las técnicas de reconocimiento de voz, conversión texto a voz y las particularidades de las plataformas a utilizar, simplificando el desarrollo de aplicaciones portables

Sin embargo, como la mayoría de lenguajes, también posee algunas desventajas, entre las que sobresalen:

- Aunque es un lenguaje que promete aportar muchos beneficios al desarrollo de aplicaciones de voz, existe cierta inseguridad aún sobre su implementación.
- Los servidores VoiceXML demandan elevadas potencialidades de procesamiento, dadas las continuas conversiones TTS y STT que se requieren.
- Para el desarrollo de aplicaciones en español existen pocas herramientas que lo soporten, ya que la mayoría están en fases de prueba o evaluación.

4. FUNCIONAMIENTO DE LOS PORTALES INTERNACIONALES MÁS SOBRESALIENTES

Tellme, Heyanita e Ydilo son 3 portales de voz reconocidos internacionalmente, los cuales han podido demostrar que las aplicaciones de voz son de gran aceptación y facilitan las actividades diarias de los seres humanos. En seguida se presenta un poco de historia, objetivos, forma de trabajo, clientes, tipos de servicios prestados y las principales soluciones desarrolladas en cada una de estas 3 empresas.

3.12 Portal de voz: TELLME

4.1.1 Reseña histórica

Tellme es el primer portal de voz. Surgió de la iniciativa de un grupo de ex trabajadores de Netscape, quienes querían que fuera posible acceder a Internet desde cualquier parte sin necesidad de contar con una PC (esto fue a mediados de 1999). El éxito del servicio sorprendió incluso a sus mismos creadores: TellMe colapsó el primer día de su puesta en funcionamiento, debido a la avalancha de llamadas que recibió.

4.1.2 Descripción general de Tellme

Tellme está permitiendo que la red telefónica comparta muchas de las mismas ideas y descubrimientos que ya hace algún tiempo hicieron grande a Internet. La meta de Tellme es acceder a Internet mediante el uso del teléfono para mejorar la interacción de las personas y negocios.

Tellme une Internet y la red telefónica debajo de una interface de usuario de lenguaje elegante para proporcionar la red de aplicación de voz más grande del mundo. Todos los días, más de dos millones de usuarios usan este tipo de aplicaciones que les permiten obtener servicio al cliente, conexión a directorios, resultados de deportes, etc.

4.1.3 Visión de Tellme

Actualmente la visión de Tellme es crear DialTone 2.0 un servicio Web personalizado para el consumidor con el objetivo de poder acceder desde el teléfono, poder interactuar con el sistema mediante la voz y realizar de esta forma cualquier tipo de actividad con solo algunas frases dichas al sistema. Ejemplo: Me gustaría comprar dos entradas para el cine.

4.1.4 Logros de Tellme

La red controlada por Tellme maneja cientos de millones de llamadas por ejemplo para Aerolíneas Delta, proporcionándoles una experiencia de servicio mejorada y permitiéndoles el ahorro de dinero. Además de reducir el costo de las inversiones de aplicaciones de voz, Tellme ha logrado construir aplicaciones que usan tecnología Web y brindarla a través del teléfono.

4.1.5 Desarrollo de los servicios

Tellme es la empresa pionera en el uso de VoiceXML, lo que le ha permitido realizar diversidad de aplicaciones que automaticen los servicios ofrecidos por muchas empresas. Habiendo manipulado la voz del usuario desde el lanzamiento de 1-800-555-TELL y AT&T VoiceDial Inalámbrico (#121), Tellme desarrolla fundamentalmente aplicaciones que permitan el aprovechamiento al máximo del uso del teléfono.

4.1.6 Descripción de las soluciones más sobresalientes que Tellme ha creado para las empresas

Tellme ha desarrollado diversidad de soluciones para las empresas consiguiendo con ello la confianza de las mismas para la automatización de sus servicios. Entre las empresas clientes de Tellme y las soluciones creadas se pueden mencionar:

- Empresa E*Trade Financiera:
Solución para consolidar la infraestructura de telefonía y redes.
La mayoría de compañías han desarrollado infraestructuras separadas de tecnología telefónica y servicios de Internet. Cada canal posee su propia lógica de negocios, equipo de hardware e integración de sistemas. La complicación y costo involucrados en la optimización llevó a esta empresa al interés de buscar otra alternativa para solucionar este problema. Fue así como optó por confiar los retos de telecomunicaciones de voz a Tellme y usar su información de Internet (ya existente) para consolidar su servicio de Internet y Telefonía. “Uniéndolo el sistema de teléfonos ya existente a la infraestructura de Internet, se pudo innovar el servicio al cliente, así como elevar las ofertas por Internet y reducir los costos de servicio al cliente”
- Fandango, la más exitosa compañía de boletaje en línea y vía telefónica de los Estados Unidos, era ya una líder en la industria con su sitio (www.fandango.com) cuando decidió ofrecer sus servicios vía telefónica. El reto de Fandango era construir un servicio telefónico de gran escala y fácil uso que fuera eficiente y económico para sus clientes. Trabajando con Tellme, Fandango se convirtió en el único proveedor de información y boletaje para el servicio 1-800-555-TELL de Tellme. A través de este compañerismo, Tellme trasladó las ofertas de formato de Internet de Fandango a VoiceXML, proveyendo así una solución a bajo costo para Fandango. Finalmente, Fandango

pudo ofrecer un servicio telefónico superior a sus clientes y ahora opera a nivel mundial a través del 1 – 888 - Fandango.

- Los proveedores de líneas telefónicas están más y más enfocados en incrementar el uso de los clientes. AT&T se unió a Tellme para proveer nuevos servicios de respuesta a voz, lanzando mLIFE para alcanzar a los clientes con nuevas ofertas. En tan solo 10 semanas, Tellme lanzó el servicio original voz #121 por mLIFE incluyendo aplicaciones para resultados de la bolsa, reportes del clima y resultados deportivos. En junio del 2003, Tellme expandió #121 con discado de voz, una aplicación que permite marcar un número verbalmente. En los primeros diez días de su lanzamiento, AT&T Wireless obtuvo 30 veces más suscriptores por medio del discado de voz de los originalmente proyectados.
- Delta es la aerolínea transatlántica líder de los Estados Unidos. Ofrece la mayor cantidad de vuelos diarios, sirviendo al mayor número de mercados y clientes, como ninguna otra aerolínea. El 15 de Abril de 2003, Delta Air Lines dio a conocer Song, un nuevo servicio desarrollado para cambiar las expectativas del cliente en cuanto a alta calidad a bajo precio. Agregado a su servicio personal, trabajando con Tellme, Song se convirtió en el primer servicio aéreo en brindar a los clientes un sistema para adquirir sus boletos por vía telefónica.

3.13 Portal de voz: HEYANITA

HeyAnita es uno de los principales proveedores de productos de software de voz. Refuerza y brinda soluciones a la industria de telecomunicaciones globales. HeyAnita proporciona un rango ancho de aplicaciones galardonadas, así como los *hosting* y servicios profesionales.

Estas soluciones permiten una comunicación con el usuario mucho más amistosa, a la vez que promueve un mayor uso de la red telefónica, con lo que reduce costos y crea nuevas alternativas para las empresas. Las aplicaciones de HeyAnita están actualmente disponibles para más de 45 millones de suscriptores a nivel mundial.

Entre sus productos se pueden mencionar:

- Voice Manager Suite: Comunicación rápida e información personal manejada por aplicaciones
- Voice Care Suite: Aplicaciones para los proveedores de Servicio
- Voice Browsing & Infotainment: Información, Entretenimiento y aplicaciones de contenido Web

4.2.1 Desarrollo de soluciones

Las soluciones que Heyanita otorga a los proveedores de diversos servicios permiten agregar valor a los productos ofrecidos a los clientes de cada empresa. Así mismo, este portal ofrece una arquitectura abierta y flexible; principalmente se apoya en especificaciones como VoiceXML 2.0 . Las aplicaciones de HeyAnita aseguran calidad y el despliegue transparente de los servicios, que integran: voz, email, marcación por voz, mensajes de texto/SMS, voicemail, e Internet.

Opciones de despliegue:

Los proveedores de servicios tienen la opción de desplegar los productos directamente a través de la red de HeyAnita o por medio del *hosting*.

El despliegue de las aplicaciones que involucra la integración de las tecnologías indispensables se realiza de forma completamente rápido permitiendo una considerable reducción en los costos y en el tiempo de comercialización.

Con la opción de *hosting* que provee HeyAnita es posible realizar ensayos de las aplicaciones, permitiendo realizar cualquier modificación de manera más rápida. HeyAnita también proporciona funcionamiento redundante, lo que evita el rango de fallos en el sistema como tal.

4.2.2 Clientes y compañías con las que HeyAnita ha trabajado

Entre los clientes de HeyAnita están:

- Sprint
- Versión Gíreles
- Ford Motor Company

Entre las compañías con las que HeyAnita a trabajado de manera conjunta están:

- Dialogic
- COMVERSE
- SpeechWorks
- Lucent Technologies
- Nuance

3.14 Portal de voz: Ydilo

4.3.1 Reseña histórica

Este portal fue fundado en 1999 por un grupo de ejecutivos con gran experiencia en varios campos de la industria y la asesoría en España. Ydilo es el primer portal de voz de habla hispana.

La especialización de esta compañía es la consultoría e implantación de soluciones basadas en reconocimiento avanzado de voz y la comercialización de este tipo de servicios en los modos “*housing*” y “*hosting*” para las empresas que así lo deseen. Los que actualmente invierten en esta compañía son: Ericsson, Mercapital, Dinamia y Corporación IBV.

4.3.2 Descripción general de las soluciones proporcionadas por Ydilo

Las soluciones que Ydilo proporciona garantizan un alto grado de automatización de los servicios de telefonía. Así mismo ofrecen una gran ventaja respecto a los call centers tradicionales, ya que desarrollan aplicaciones que utilizan el reconocimiento de voz, lo que permite que los sistemas sean más agradables al usuario.

Las empresas que contratan los servicios que Ydilo proporciona han podido observar considerables ventajas competitivas, crecimiento en la satisfacción de los clientes, reducción de costos, mejoramiento de la seguridad y lanzamiento más ágil de nuevos productos.

4.3.3 Objetivos de Ydilo

Debido a la importancia de distribuir el contenido al público a través de la interfaz más adecuada, Ydilo se ha propuesto ofrecer una serie de beneficios clave a todas las empresas que contraten sus servicios, entre los que se encuentran:

1. Satisfacción del cliente

Las aplicaciones de voz diseñadas por Ydilo certifican una experiencia rápida y agradable para el acceso a la información. Todo ello debido al reconocimiento de voz, el cual elimina las esperas tan tediosas en la espera de un receptor humano o el acceso mediante menús activados por tonos.

2. Reducción de costos.

Por lo regular la mayoría de los call centers prestan el servicio mediante menús IVR, lo que involucra mayores costos. Sin embargo, Ydilo presenta la opción de automatizar este servicio utilizando sus aplicaciones de voz, lo que reduce considerablemente los costos, ya que la duración de las llamadas es mucho menor pues los menús son más rápidos y naturales.

3. Aumento de ingresos.

Todas las transacciones que se puedan efectuar mediante la voz tienen una gran cualidad: poseen un alto número de accesos utilizando el teléfono, por lo que es fácil deducir que estos sistemas de voz son rápidamente aceptados, lo que garantiza a las empresas oportunidades de realizar mayores ingresos.

4. Mayor seguridad

Las típicas preguntas de seguridad, como el número de PIN, entre otros, es más sencillo hacerlas mediante los sistemas que poseen reconocimiento de la voz. Las tecnologías que verifican al hablante se encargan de examinar las características vocales de las palabras habladas (huella vocal). Todo esto hace que el cliente ya no tenga que recurrir al proceso largo de los sistemas IVR.

5. Acelera la salida al mercado

Utilizar la plataforma a nivel de operadora de Ydilo asegura a los clientes el rápido despliegue de servicios de voz de alta calidad y con el mínimo coste de establecimiento

6. Aumento de la competitividad

Las empresas que deciden utilizar los servicios de voz de Ydilo, además de innovar sus servicios, demuestran un liderazgo en su compromiso con la satisfacción del cliente, ya que en la mayoría de industrias los sistemas utilizados son de marcación activada por tonos, por lo que de esta forma las empresas pueden diferenciarse de la competencia y, a la vez, mantener sus clientes.

4.3.4 El trabajo con los clientes

El equipo de proyectos de Ydilo trabaja con clientes de diferentes industrias para los cuales ha desarrollado diversidad de servicios de voz como: venta de entradas, banca telefónica, información de vuelos y pedidos, entre otras. Al realizar el diseño de cada servicio de voz, Ydilo se compromete a examinar cuidadosamente las estrategias que funcionan y las que no, así como a conseguir mayores niveles de satisfacción por parte del usuario. Para poder realizar una solución que verdaderamente cumpla sus expectativas, el personal de Ydilo que trabaja con el cliente se encarga de identificar cuál sería el mejor uso de los servicios de reconocimiento de voz que contribuirían al mejoramiento de la empresa cliente, y también examina la posible integración con los procesos internos y sistemas heredados. “Ydilo ayuda a las empresas a entender las posibilidades y limitaciones de los sistemas de voz para obtener los máximos beneficios”⁷.

4.3.5 Diseño y desarrollo de los servicios

Ydilo se ha concentrado en el diseño de alta calidad de los diálogos. Al decir diseño de un diálogo se están incluyendo los servicios que se ofrecen al cliente, el desarrollo del diálogo de las llamadas (el cual es realizado por lingüistas expertos) y la elaboración de las rutinas de recuperación de errores, lo que permite asegurar la fluidez en las llamadas.

Después de haber diseñado el diálogo de las llamadas, se comienza a codificar la aplicación en Java y VoiceXML. Para poder agilizar el tiempo de salida al mercado de los sistemas, Ydilo ha construido aplicaciones reutilizables (diálogos y módulos que se pueden desplegar como parte del proyecto de un cliente).

Las bases de datos existentes, las aplicaciones heredadas, los call centers e incluso las soluciones IVR ya desplegadas pueden integrarse para conseguir mejores resultados y elevar el nivel de prestaciones de la plataforma para poder satisfacer futuras necesidades.

Debido a que todo lo referente al audio debe ser agradable y de óptima calidad, Ydilo contrata locutores profesionales y cuenta con sus propios estudios de grabación y postproducción para alcanzar una calidad perfecta no solo en “personalidad”, sino en ingeniería del sonido.

Ydilo ha observado que no es suficiente con realizar diseños de alta calidad en los diálogos, sino que las pruebas y ajustes con los usuarios son claves para que los servicios de voz sean satisfactorios y alcancen los objetivos del cliente. La forma como Ydilo realiza las pruebas es: en primer lugar, afina y prueba la aplicación durante 6 semanas (para tener justo lo que el cliente desea). Luego son transcritas las llamadas de los usuarios reales con el fin de encontrar los puntos donde los usuarios pueden tener dificultades al usar el servicio, pues son la clave para que los servicios de voz sean satisfactorios y alcancen los objetivos de los clientes. La dinámica de Ydilo es afinar y probar una aplicación a lo largo de unas seis semanas para llegar a conseguir los niveles deseados de satisfacción del usuario.

4.3.6 Infraestructura de Ydilo

Con el fin de poder proporcionar un servicio que cuente con el 99.9% de disponibilidad, así como garantizar plena seguridad en las aplicaciones y en los datos de los clientes, Ydilo cuenta con lo siguiente:

- Acuerdos de confidencialidad por parte de los empleados
- Tarjetas de acceso y escáneres de huellas dactilares para el control del acceso a las oficinas de Ydilo
- Guardias de Seguridad (durante las 24 horas)
- Firewalls de última generación y programas de antivirus profesionales para la protección del servidor y la Intranet del intrusismo.
- Los servidores ASR, TTS, PBX, bases de datos y servidores de aplicaciones están instalados en ordenadores físicamente diferentes
- Planes de contingencia para una rápida recuperación en caso de fallo del sistema
- Todos los sistemas son controlados por un equipo técnico especializado, durante las 24 horas del día los 365 días del año
- Una línea de información está disponible para los clientes las 24h
- El centro de datos está equipado con un suministro eléctrico ininterrumpido y unos avanzados sistemas contra incendios
- Conexiones redundantes de telecomunicaciones e Internet contratadas a diferentes operadoras de telecomunicaciones para asegurar la imperceptible reanudación del servicio en caso de fallo del sistema
- El HW y SW que Ydilo utiliza para los servicios son suministrados por las empresas líderes del mercado
- Todas las aplicaciones están monitorizadas, lo que permite realizar informes acerca del funcionamiento del sistema. Los clientes pueden acceder a dichos informes en tiempo real, mediante una extranet

Los informes contienen información sobre:

- Tráfico de llamadas
- Perfil del usuario (basado en el origen de los números de teléfono)
- Uso de la aplicación
- Acceso al contenido
- Eficacia del reconocimiento y utilización de la gramática
- Diálogos completos del usuario

4.3.7 Diversidad de soluciones ofrecidas por Ydilo según el sector

• Soluciones para los *call center*

Actualmente este servicio se ha convertido en una necesidad, principalmente para aquellas empresas que requieren estar disponibles las 24 horas al día durante el año. Se basa en sistemas interactivos de voz (IVR), lo que requiere una constante gestión de agentes. A cambio de esto, las soluciones basadas en reconocimiento de voz que ofrece Ydilo aumentan la eficacia de los call centers, reducen sus costos y aumentan la satisfacción de sus usuarios. Al automatizar estos sistemas, los agentes pueden concentrarse en actividades de mayor valor añadido para la empresa, se elimina la posibilidad de error humano, se ofrece mayor seguridad en los contenidos de las bases de datos, se aumentan las ventas, y se facilita el mantenimiento de las BD.

• Soluciones para la administración pública

Son muchos los servicios que se pueden brindar mediante un portal de voz. Ydilo presenta también opciones para organismos gubernamentales de ámbito nacional, para los cuales se pueden desarrollar sistemas que permitan difundir información como:

- Horarios de atención al público, direcciones y rutas de acceso
- Servicios de petición de cita

- Ofertas de empleo
- Becas
- Subvenciones a empresas
- Recomendaciones de seguridad para viajeros
- Instrucciones sobre actuación en situaciones de emergencia
- Procedimientos y políticas de la Administración
- Pago de multas e impuestos varios
- La operadora virtual (actúa igual que el mostrador de recepción y conecta automáticamente con personas o departamentos específicos).
- Captación de opiniones de los ciudadanos
- Servicio de notificación personalizada (recuerda las citas)

La implementación de este tipo de soluciones aprovecha las inversiones previas en HW y SW, lo que permite un mejor aprovechamiento de los fondos públicos, a la vez que da la opción a que un mayor porcentaje de la población pueda tener acceso a Internet. (Un ejemplo de esto es que en España el 30% de la población tiene acceso a Internet mientras que un 98% tiene acceso a una línea telefónica).

• **Soluciones para transporte y logística**

La informática se ha convertido en una parte importante de las compañías de transporte, pues con la misma ellos pueden ejercer un mayor control sobre sus servicios, como por ejemplo: pueden conocer si su envío ha sido entregado y, si no ha sido entregado, la razón, o también pueden saber si un conductor puede o no completar la ruta planeada. Ahora bien, con los servicios de voz se pueden crear interfaces mucho más sencillas y eficientes tanto para los empleados como para los clientes.

Entre las aplicaciones de voz para la industria del transporte están:

- Función de seguimiento de las entregas por voz
- Situación del inventario e información del envío
- Información de la entrega
- Incidencias y excepciones en la manipulación
- Operaciones de recogida y embalaje guiadas por voz
- Recogida y entrega de talones
- Comprobación de inventario, recuento periódico
- Notificaciones por voz sobre el envío

Las interfaces que desarrolla Ydilo, además de ser de manejo sencillo y disponibilidad permanente, permiten introducir los datos y las incidencias directamente en la base de datos, desde cualquier teléfono y solamente hablando.

• Soluciones para viajes y turismo

Para las empresas dedicadas a viajes y turismo, Ydilo ofrece automatizar la mayoría de las consabidas peticiones de los clientes. Los servicios de voz pueden incluir:

- Información en tiempo real sobre salidas y llegadas
- Información sobre destinos y horarios
- Notificaciones personalizadas acerca de posibles cancelaciones y retrasos
- Situación de la entrega
- Lugar de recogida del pedido
- Venta de billetes de viaje

- **Soluciones para ventas al por menor**

Este tipo de soluciones sirven especialmente para la venta de productos de uso común que necesitan pocas explicaciones o para pedir productos anunciados en la TV, la radio o en medios impresos.

Los servicios de voz para empresas de venta al por menor pueden incluir:

- Venta por teléfono (permite al cliente solicitar información sobre la empresa y sus productos, material informativo así como adquirir productos y servicios; el cliente puede acceder al estado de su cuenta personal y realizar pagos utilizando un servidor seguro)
- El servicio de alerta (mantiene a su cliente al corriente del estado de su pedido, le informa de las ofertas existentes basándose en su historial de pedidos y/o lugar de residencia, y todo ello costando una mínima parte de lo que costaría emplear un teleoperador humano para realizar el servicio)
- Localizador de sucursal/oficina/tienda (permite al cliente encontrar fácilmente la oficina o tienda más cercana, conocer sus horarios y el modo de llegar al establecimiento, y si lo desean les pone directamente en contacto con el establecimiento)
- Las operadoras virtuales (actúan como el mostrador de información de su empresa conectándole directamente con las personas o departamentos específicos)

4.3.8 Breve descripción de las soluciones más sobresalientes que Ydilo ha desarrollado para empresas y eventos

Ydilo cuenta con una gran aceptación por parte de diversas empresas no solo en España sino internacionalmente. Entre las soluciones que Ydilo ha desarrollado para las empresas y los eventos en los que ha participado están:

- Para la empresa: INTIX, Ydilo presentó los servicios de reconocimiento de voz para la venta de entradas en “INTIX in Denver 2003”, del 21 al 24 de enero de 2003.
- Ydilo fue uno de los principales patrocinadores del "Call Center + CRM Solutions 2002" en Madrid, del 19 al 21 de noviembre de 2002, en donde se realizó una presentación sobre el reconocimiento de voz en el mundo real.
- VAD es uno de los eventos más importantes dedicados a la venta a distancia en Francia. Ydilo estuvo en la exposición, donde pudo mostrar cómo el reconocimiento de voz puede facilitar y optimizar la venta por catálogo, radio o televisión.
- Ydilo participó en Madrid en un evento organizado por Callware e Intel el 24 de octubre de 2002.
- "V-World Europe", del 10 y 11 de septiembre de 2002, es la conferencia anual de Nuance Communications, una de las principales empresas en la industria del Reconocimiento de Voz. Ydilo presentó sus estudios más recientes que muestran el funcionamiento de los servicios de voz en el mundo real.
- Ydilo fue copatrocinador del "Global Outsourcing 2002", en Madrid, el 27 de junio de 2002. Allí se mostró cómo se pueden reducir los costes, a la vez que se mejora la atención al cliente, utilizando el reconocimiento de voz como un servicio subcontratado.
- Ydilo estuvo presente en el foro de "Mobile Internet Services 2002", en Ginebra, Suiza, el 17 de junio de 2002.

- Ydilo estuvo presente en el foro para usuarios que se celebró en Madrid sobre "Information Builders", el 13 y 14 de junio de 2002 e hizo una presentación sobre la integración de interfaces de voz basados en bases de datos.
- Ydilo estuvo presente en la conferencia "SpeechTEK 2001" que se celebró en Nueva York/EE.UU, del 29 de octubre al 1 de noviembre de 2001.

Informes más destacados de Ydilo durante el año 2003

- En abril de 2003 la solución Ydilo Tráfico para DGT atendió más de 250.000 llamadas en 10 días durante situaciones críticas.
- Luego, en el mes de mayo del mismo año, Ydilo fue reconocida entre las 3 primeras empresas del mundo en innovación por servicios de reconocimiento de voz.
- En diciembre de 2003, Vodafone España informó que instalará un nuevo sistema automático de voz para servicios de atención a clientes, concebido y operado por Ydilo.

Mediante la descripción del funcionamiento y los servicios que ofrecen estos 3 portales, se pudo observar la diversidad de usuarios que se pueden captar, así como también las grandes ventajas que las empresas tendrían al presentar información a través de estos portales. Estas son las razones por las que se puede decir que en el futuro, cuando todas las tecnologías involucradas estén al alcance (económicamente hablando) también en los países en vías de desarrollo, los portales de voz se convertirán no solo en una gran ayuda para la mayoría de la población, sino en una necesidad.

5. EXPECTATIVAS PARA EL FUTURO

5.7 Fuerzas que demandan la Web de voz

La cantidad existente de teléfonos tanto fijos como móviles que poseen y utilizan las personas es relativamente alta frente al número de ordenadores conectados a Internet. Recuérdese que:

- a) Son limitadas las posibilidades de los teléfonos móviles para el acceso modo texto a los contenidos de Internet: pantalla pequeña, reducida memoria, reducida capacidad de procesamiento, bajas velocidades de transmisión.

- b) Las demanda de acceso de los contenidos de Internet desde dispositivos de uso común por la mayoría de personas (teléfonos).

Estas son las principales fuerzas que permitirán que en el futuro los portales de voz puedan constituirse en uno de los servicios de mayor uso en la mayoría de los países.

5.8 Generalidades y expectativas del futuro de los portales de voz

5.8.1 Generalidades

Los portales de voz actualmente son conocidos por ofrecer los servicios de: noticias, tiempo, bolsa, deportes, lectura de correo electrónico, entre otros. Sin embargo, se espera que en el futuro se pueda realizar una navegación por la totalidad de la red a través de la voz.

Se sabe que el número de personas que acceden a Internet crece a un ritmo mayor del 100% anual; así mismo, la telefonía móvil lo hace a un ritmo entre el 60%-80%. A esta información se une el hecho de que los negocios por Internet cada día aumentan, por lo que se espera que en el futuro se consolide como el centro de negocios, y por lo tanto, pase a ocupar un lugar importante en la base de la economía del nuevo milenio. De tal manera que los portales de voz por estar dentro de este ambiente poseen un futuro prometedor. Por ello las operadoras telefónicas, los fabricantes de terminales, los portales tradicionales, las empresas proveedoras de contenidos o las grandes marcas, ven en los portales de voz una gran oportunidad para hacer negocios y una razón para aliarse con otras compañías y hacer inversiones conjuntas

5.8.2 Expectativas de los portales de voz

Los portales de voz esperan que la fuente de ingresos sea el cobro de comisiones por las compras que realicen los usuarios a través de su servicio.

Al referirnos un poco más a cifras se puede decir que las aplicaciones de voz, o el software de reconocimiento del habla, pasarán de un volumen de negocio de 200 millones de dólares a 2,700 millones, aproximadamente, al final de 2005. Así mismo, se espera que las transacciones y la publicidad asociada a los portales de voz lleguen a mover 5,000 millones de dólares en 2005. Más otros 6,000 millones de dólares en concepto de hardware asociado, software y proveedores de servicios en red.

En cuanto a usuarios se puede mencionar que, según el informe elaborado por The Kelsey Group, se estima que para 2005 habrá 18 millones de usuarios de algún tipo de portal de voz.

5.8.3 Perfil de los desarrolladores de los portales de voz

Lo interesante del desarrollo de los portales de voz y todo tipo de aplicaciones de voz es que las personas que se dediquen a ello no necesitarán ser expertos en el tema, ya que el lenguaje VoiceXML cuenta con el grado de facilidad necesario para que cualquier programador de otro lenguaje pueda comprenderlo rápidamente. Así mismo, todo el resto de herramientas que ya existen incorporan los procesos complejos de manipulación de la voz. Esto permitirá que las empresas no necesiten especialistas en el desarrollo de sistemas de voz, pues las herramientas están en el mercado, por lo que lo único que se requerirá será un poco de entrenamiento, práctica y creatividad. Uniendo estos tres elementos, el desarrollar una aplicación de voz será muy sencillo.

5.9 El futuro del desarrollo tecnológico

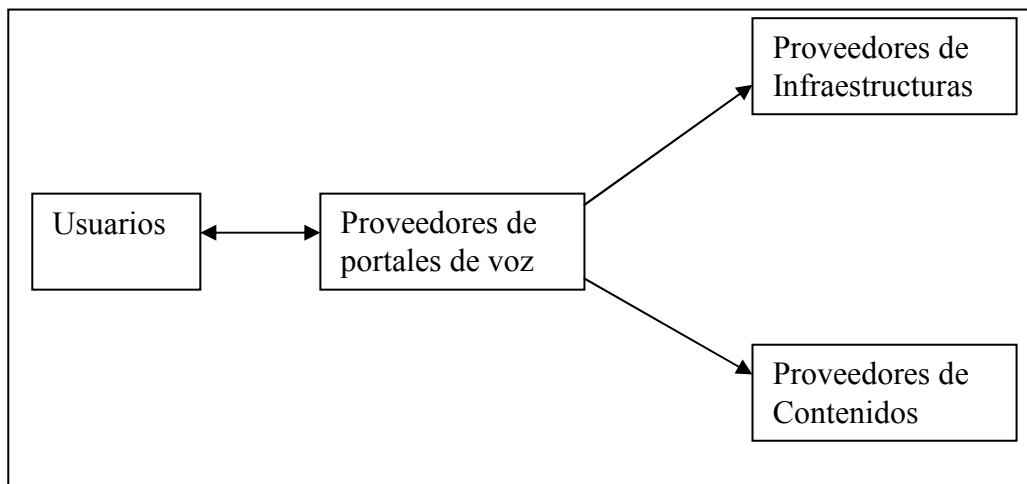
5.9.1 Desarrollo tecnológico

La introducción que durante los años 90 tuvieron, comercialmente hablando, las aplicaciones de reconocimiento de voz, así como los servidores de red capaces de soportar dichas aplicaciones, el procesamiento de voz, los navegadores de voz y los lenguajes apropiados para la manipulación de la voz, hizo posible el despliegue de una serie de tecnologías relacionadas como la tecnología TTS, tecnología de autenticación de voz entre otras. Al mismo tiempo, dentro del desarrollo tecnológico telefónico también se ha observado el surgimiento de nuevas tecnologías, tal es el caso de GPRS (General Packet Radio Services), el cual es un conjunto nuevo de servicios los cuales se añadirán a los actuales que posee GSM (Global System for Mobile Communications, lo que en español significa Sistema Global para Comunicaciones Globales, un servicio universal que permite utilizar el mismo número y teléfono en todos los países del mundo que poseen la misma tecnología).

Básicamente, así se añade conmutación de paquetes de datos a todos los niveles de la red GSM. GPRS ofrece funciones de autenticación, control de accesos, confidencialidad de la identidad del usuario y confidencialidad de la información. Otra tecnología reciente pero muy prometedora es UMTS (Universal Mobile Telephone System) sistema de telefonía móvil de 3ª generación. UMTS será capaz de alcanzar velocidades entre 384 kbps, para redes de banda ancha y 2.0 Mbps para locales. En el futuro se espera incorporar estas tecnologías al servicio de los portales de voz, lo que permitirá que con GPRS, y sobre todo con UMTS, los usuarios sean capaces de obtener información contra un terminal, en otras palabras introducir y obtener datos, mediante voz, de maneras mas rápidas, seguras y confiables, permitiendo de esta forma relevar por completo, para este servicio, los teclados.

5.9.2 Protagonistas en la Web de voz

Figura 13. Protagonistas en la Web de voz



Usuarios:

Cualesquiera individuos que deseen utilizar el servicio.

De estos depende en gran parte el progreso y la incorporación de nuevos servicios a los portales de voz, ya que ellos serán los que harán uso de los mismos, por lo que el avance de estas aplicaciones de voz será proporcional a la frecuencia, importancia y utilidad que le den al servicio.

Proveedores de portales de voz

Compañías que ofrecen o planean comercializar servicios de portales de Voz, por ejemplo:

- portales WEB
- proveedores de Servicios de Red

Debido a que ya conocen parte del mercado potencial de este servicio, son los encargados de realizar la publicidad correspondiente, permitiendo que esta sea atractiva y ofrezca todo lo que los clientes o usuarios desean, principalmente en aspectos de comodidad, disponibilidad y seguridad.

Proveedores de contenidos

Compañías de servicios de información, de comercio electrónico, etc., interesadas en expandir sus servicios a bajo costo. El papel fundamental de estas compañías es habilitar sus informaciones no solo para la Web tradicional sino para voz esto les permitirá venderlas a los proveedores de portales de voz.

Proveedores de Infraestructuras

Proveedores de tecnologías SW (reconocimiento de voz, identificación de localización), compañías de hosting, desarrolladores de Web (servidores de aplicaciones Web, etc.), proveedores de plataformas y HW, proveedores de transporte de red. Estos proveedores se convertirán en el puente imprescindible para poner en marcha los portales de voz.

5.10 Los portales de voz y el comercio electrónico

El comercio electrónico llamado también online es otro de los destinos que esperan a los portales de voz. Sin embargo aun es necesario confirmar aspectos relacionados con la seguridad de las transacciones comerciales que para este caso no solo serían online sino orales.

5.10.1 V-Commerce

E-Commerce, término utilizado para referirse al comercio electrónico para este tiempo es ya una expresión popular. Sin embargo, debido a los constantes avances tecnológicos, nuevas modalidades de realizar las transacciones han surgido. Prueba de ello es el M-Commerce, representación de la actividad comercial realizada mediante servicios específicamente móviles inalámbricos. En su mayoría este tipo de actividades facilitan el monitoreo de vehículos, equipos, bienes y procesos. Así mismo, permiten al usuario realizar actividades sencillas, como consultar el tráfico o comprar algún artículo, todo ello en su mayoría interactuando con el teclado de su móvil.

Ahora bien, V-Commerce es el nuevo término que permitirá realizar las transacciones mediante voz, para lo cual se espera que las tecnologías de voz puedan ser capaces de reconocer a sus interlocutores. La empresa Nuance (Empresa líder en soluciones de voz) dirige el consorcio de más de 25 empresas dedicadas a establecer un estándar de seguridad para los sitios de comercio on-line por voz.

5.10.2 Seguridad

La seguridad es un elemento fundamental al realizar transacciones comerciales (online principalmente) por brindar la confianza que toda persona necesita. Al estar este elemento confirmado, se espera que dé lugar a nuevos esquemas de negocios.

Actualmente el trabajo en este aspecto es bastante arduo. A pesar de ello, el desarrollo de los servicios transaccionales ha crecido (principalmente en los países desarrollados). Algunos de los sistemas de verificación de voz permiten identificar a las personas con un factor de seguridad diez veces superior al de la huella digital. Esto, junto con una clave operativa, permite que las transacciones sean prácticamente seguras. En Estados Unidos, compañías como Home Shopping Network utilizan ya esta tecnología de autenticación en sus transacciones.

Cuando la mayoría de sistemas de voz cuenten con el grado de seguridad aceptable, será factible realizar las compras de última hora con solo tener un teléfono de cualquier tipo. Esto concordará completamente con el mundo tan agitado en que vivimos, donde por disponer cada vez de menos tiempo se desea optimizarlo.

5.11 Desarrollo y expectativas de portales de voz en Guatemala

5.11.1 Viabilidad técnica, económica y social

Es importante conocer la posibilidad que se tiene para poner en marcha el servicio de portales de voz en Guatemala, tanto técnica, económica y social. Así se determinará si es factible implementar este tipo de soluciones en nuestro país a corto, mediano o largo plazo.

5.11.1.1 Técnica

Como se ha podido observar en los capítulos anteriores para poder desarrollar y poner en marcha aplicaciones de voz es necesario contar con lo siguiente:

En primer lugar, una plataforma telefónica para que el acceso pueda ser mediante cualquier tipo de teléfono (ya sea convencional o móvil), también se requiere de hardware que soporte la tecnología asociada a este servicio. Para la parte de desarrollo de los sistemas se requieren herramientas de reconocimiento de voz, de texto a habla (TTS) (en este caso este tipo de herramientas deben estar en español), así como editores de VoiceXML o similares.

En Guatemala se tiene la dicha de contar con 4 empresas telefónicas que brindan sus servicios; la mayoría de ellas ofrecen tecnologías modernas. Por ello el primer peldaño está superado. Siguiendo los requerimientos necesarios se puede notar que el HW tampoco es obstáculo, puesto que existe un gran número de proveedores de HW, y si cualquier componente aún no se encuentra en el país, existe la opción de traerlo de otro país donde sí exista, comercialmente hablando. Con respecto a las herramientas de desarrollo, también pueden ser adquiridas. Claro está que por lo general tendrían que adquirirse en el exterior. Sin embargo, existen y por lo tanto dejan de ser obstáculo técnico (Entre algunas herramientas TTS en español están Nuance V-Builder). Los interesados en probar estas herramientas pueden utilizar simuladores de BeVocal o Tellme.

Todos los componentes involucrados en este nuevo servicio existen; pese a ello es preciso integrarlos, para lo cual la empresa que desee hacerlo debe tener presente que, aparte de la fuerte inversión que debe realizar, para que el servicio llegue al mayor número de usuarios posibles y pueda recuperar la inversión prontamente se necesita lanzar una fuerte campaña publicitaria, pues está comprobado que al motivar a los guatemaltecos a utilizar un servicio (un ejemplo de ello es “Telepromos de Guatemala”) el número de usuarios crece considerablemente.

5.11.1.2 Económica

La viabilidad económica principalmente evalúa el costo que requiere toda la tecnología implicada en el desarrollo de un portal de voz, así como el costo del servicio para los usuarios.

No es posible describir el costo de cada elemento incorporado en este tipo de servicio (debido a ciertas reglas de privacidad de las empresas que la poseen), pero, como se sabe, toda tecnología novedosa tiene un costo alto. Sin embargo, conforme el tiempo avanza y se descubren nuevas tecnologías, las actuales van siendo más accesibles, por lo que se espera que en poco tiempo todas las tecnologías indispensables para este servicio se puedan adquirir sin mayores complicaciones económicas.

5.11.1.3 Social

Quizás sea este el principal obstáculo para el desarrollo de los portales de voz. Por un lado, existe el temor en las compañías en invertir capital en un servicio que en la actualidad es prescindible y que cubriría un segmento de mercado bastante pequeño, ya que las personas con capacidades de costearlo son, en relación con la población total de Guatemala escasas.

Por otro lado, el mismo nivel de desarrollo del país crea un ambiente poco llamativo, ya que la población en general se preocupa principalmente por cubrir sus necesidades básicas. Dadas estas circunstancias, sería necesario lanzar una campaña publicitaria muy fuerte e incluir diversos servicios, no solo los comunes (noticias, deportes, estado del tiempo, entretenimiento, tráfico), sino variedad de servicios que incluyeran las actividades molestas pero obligatorias, como las grandes colas para el pago de servicios públicos. Si se tomara en cuenta una visión más amplia, sería posible encontrar un mercado mayor para los portales de voz y, por ende, una inversión muy conveniente.

Luego de haber descrito las viabilidades técnica, económica y social, se puede concluir que si alguna compañía desea lanzarse a realizar un negocio de este tipo a corto plazo necesita involucrar variedad de servicios y una adecuada publicidad para poder abarcar un buen número de usuarios. Así mismo, debe realizar cualquier tipo de alianza con alguna de las empresas de telefonía que ya existen. En el siguiente segmento se mostrará un ejemplo de una empresa que ha comenzado a prestar este servicio en Guatemala.

5.12 Primer portal de voz en Guatemala

Una de las empresas que brinda el servicio de telefonía en Guatemala ha decidido crear un portal de voz. Obviamente, solo las personas que poseen teléfono (tanto residencial como móvil) de esta empresa pueden acceder a este servicio.

La empresa que proporciona el servicio es BellSouth y el nombre del portal de voz es “Internet Móvil Voz”. Este nombre alude al hecho de poder tener acceso a Internet desde cualquier teléfono.

Internet móvil voz:

Este servicio fue creado para ayudar a volver más práctica la vida diaria de los clientes. Es un servicio que permite por medio de comandos hablados, acceder a productos y servicios que en su mayoría provienen de Internet.

El objetivo es brindar acceso a Internet mediante la telefonía móvil, sin necesidad de poseer un computador y desde cualquier parte donde el usuario se encuentre.

Una característica muy importante de este servicio es que no hay necesidad de ver la pantalla para visualizar datos, ni utilizar las teclas, ya que toda la información se proporciona exclusivamente en forma hablada.

Acceso a BellSouth Internet móvil voz:

Para acceder a este nuevo servicio solo es necesario marcar *999 inmediatamente se podrá acceder a una amplia gama de información de actualidad, y a variados contenidos de interés.

Información contenida en este portal de voz:

El menú principal de este portal tiene las siguientes opciones:

- Noticias
- Deportes
- Negocios
- Entretenimiento
- Cartelera de cines
- Astrología
- Humor

- Clima
- Correo electrónico
- Agenda

Los servicios de correo electrónico y agenda se encuentran disponibles solo para los usuarios que deseen activar el servicio Oficina Móvil (el cual se explicará mas adelante).

Interacción con el portal de voz

Para poder encontrar la información específica de cualquiera de estas opciones únicamente se debe repetir el nombre de la opción. Luego de ello el portal le dirá un submenú con las opciones referentes a la opción del menú principal que eligió. Por ejemplo: noticias: nacionales o internacionales, deportes: fútbol nacional e internacional; astrología: horóscopo; entretenimiento, cines, restaurantes etc. Cuando se llega a la información requerida en algunos casos se ofrecen detalles de la misma, por ejemplo en el caso de una película, ya que si el usuario desea tener la idea general sobre el contenido de la película puede solicitarlos.

Comandos principales de Internet móvil voz

En cualquiera de las secciones donde el usuario se encuentre puede recurrir a decir uno de estos comandos: “menú principal”, “ayuda”, “dónde estoy”, “hasta luego”. Con estos comandos el usuario puede sentir que la interacción con el portal es mucho más sencilla.

Si en algún momento el usuario no recuerda la sección en la que está, simplemente puede recurrir a decir “¿Dónde estoy?” y el sistema le recordará la sección donde se encuentra.

Con una misma llamada es posible ir a diversas secciones; solo se debe decir menú principal y desde allí escoger la sección deseada.

Cuando el sistema no entiende lo que el usuario dice, es debido a una mala conexión telefónica o a mucho ruido de fondo lo cual dificulta la comprensión. En estos casos el sistema le solicita al usuario que repita lo dicho. Sin embargo, si aun así el sistema no comprende, es mejor realizar nuevamente la llamada.

El servicio siempre se podrá tener desde cualquier teléfono BellSouth, no obstante, los servicios estarán personalizados para el dueño del teléfono.

Para salir del sistema, el usuario puede decir "Hasta luego" o simplemente cortar la llamada.

Oficina móvil

Es un producto cuyo objetivo es simplificar la ocupada vida personal y profesional de los usuarios. El servicio Incluye:

1. Una cuenta de correo electrónico con acceso desde el teléfono mediante comandos de voz
2. Notificaciones vía mensajes de texto cuando se recibe un nuevo correo electrónico
3. Agenda WEB sincronizable con el sistema de correo (Outlook) o con la Palm. Adicionalmente este sistema realiza el envío de recordatorios vía mensajes de texto, antes de las citas
4. Servicio de mensajería instantánea, el cual permite conectarse simultáneamente y en tiempo real con los servicios más populares de mensajería existentes en la actualidad

5. Versión personalizada del portal “BellSouth Internet Móvil” lo que permite al usuario seleccionar el contenido que desee, disponiendo de él cada vez que ingresa a <http://www.bellsouthim.net.gt/>
6. Servicio de mensajes de contenido informativo de actualidad como: noticias, tipo de cambio, estrenos cinematográficos, por citar algunos; los cuales son enviados directamente al teléfono a través de mensajes de texto
 - Servicio de correo electrónico:
Al tener este servicio activado, el usuario puede no solo leer sus correos desde su computadora, sino escucharlos a través del portal de voz desde su teléfono. Así mismo, si el usuario lo desea puede configurar su cuenta para recibir alertas en el teléfono acerca de los mensajes que le interesen.
Este servicio le ofrece al usuario 3 megas de capacidad para recibir e-mails de sus otras direcciones de correo electrónico en su e-mail de Oficina Móvil.
Igualmente es posible responder a los e-mails mediante la voz, llamando siempre al *999
 - Servicio de agenda móvil:
Este servicio le permite al usuario tener siempre a mano toda la información personal y de negocios que desee, con lo que se logra organizar mejor el tiempo y se evita olvidar alguna cita o teléfono importante. También es posible sincronizarla con la Palm o la PC.
El contar con este servicio le permite al usuario tener siempre acceso a su listado de direcciones ya que puede consultarlo desde cualquier computadora o desde su teléfono. La información personal que maneja este servicio nunca es compartida lo que permite brindarle al usuario completa confidencialidad.

Costo del servicio:

El servicio está disponible para todos los clientes BellSouth a nivel nacional. El costo por minuto es de Q1.50. Para los usuarios que además deseen adquirir el servicio de Oficina Móvil el costo es de Q15.00 mensual⁸.

Así pues como se puede observar que en nuestro país ya se cuenta con lo necesario para comenzar a incorporar este nuevo tipo de servicios. Sin embargo esta tecnología aún esta en sus inicios, por lo que hasta dentro de unos años la mayoría de personas notará su presencia y recurrirá a la misma.

CONCLUSIONES

1. En la actualidad, la necesidad de encontrar cualquier tipo de información en el menor tiempo posible y con el menor uso de recursos, ha llevado al descubrimiento de tecnologías que hagan este hecho una realidad. Internet se ha situado como el centro proveedor de información, y día a día se trabaja en el desarrollo de nuevos canales para poder acceder a dicha red. En su tiempo las páginas estáticas proveyeron la información deseada; luego fueron las dinámicas, ahora lo es WAP conjuntamente los portales de voz, los cuales permiten acceder a Internet desde un teléfono convencional o móvil.
2. Los portales de voz hacen posible el establecimiento de la convergencia del mundo Internet con el mundo de la telefonía vocal más elemental, lo cual permite el acceso no solo a los clientes del servicio telefónico fijo, sino también a los del servicio móvil. Es importante comentar que este nuevo canal de acceso a Internet no reemplaza a los actuales navegadores visuales utilizados por la mayoría de personas, sino que su finalidad es abrir un nuevo medio de servicios online a los que se pueda acceder más fácilmente, ya que la interacción con el sistema a través de este medio es la voz, dando realce de esta forma al pensamiento de muchos en el sentido de que es mejor hablar que escribir. Sin embargo, también es necesario aclarar que debido a la juventud de esta tecnología, además de presentar ventajas también se enfrenta a diversos problemas, principalmente relacionados con la dificultad de que el sistema entienda correctamente lo que están diciendo los usuarios, por tal razón las investigaciones para mejorar el reconocimiento de voz continúan perfeccionando sus técnicas y procedimientos.

3. El lenguaje VoiceXML permite realizar la descripción de diálogos únicamente vocales. En los mismos se define qué debe decir la máquina que está interpretando la página y cuando debe esperar una entrada de datos del usuario (ya sea por reconocimiento de voz o por detección de tonos multifrecuencia). La navegación en este tipo de páginas es muy sencilla, ya que cada página conduce a otra mediante la interacción que el usuario tenga a través de la voz. La estandarización de VoiceXML dará como resultado una enorme popularización de los servicios interactivos de voz, básicamente por la sencillez del lenguaje (similar a html), lo que permitirá que los proveedores de contenidos no tengan que disponer de especialistas en el desarrollo de sistemas de voz. Así mismo, el hecho de que, con un esfuerzo mínimo de desarrollo, cualquier dispositivo pueda generar páginas VoiceXML, abre nuevas posibilidades en los campos de la telemonitorización y la domótica, con lo que se lograría que cualquier usuario pudiera oír y hablar con los elementos monitorizados.

4. Con la diversidad de aplicaciones vocales que ofrecen los portales de voz reconocidos internacionalmente es posible deducir, que en cualquier caso, parece claro que la voz, como interfaz, tiene un éxito asegurado. Y permite augurar que la voz será en un futuro la que dirija la relación del usuario con Internet. Además, las soluciones implementadas para automatizar servicios ya existentes dan como resultado la eliminación del fallo humano, asegurando exactitud en los contenidos incorporados a las BD, lo que permite el incremento en las ventas, a la vez que facilita el mantenimiento de las BD, y minimiza costos.

5. La demanda de este servicio en Guatemala es aún pequeña, debido principalmente a diversos factores socioeconómicos que dificultan el progreso de este tipo de tecnología. Sin embargo, ya se cuenta con un portal de voz, el cual proporciona información básica que puede ser escuchada desde un teléfono.

6. Se espera que en el futuro el interés por poseer una audioWeb sea similar al que hoy existe por contar con una página Web. Esto permitirá que el despliegue de servicios sea más eficiente, cómodo, fácil y económico. Así mismo, con el avance en el desarrollo de la integración de tecnologías se espera que, a partir de una página html o wml se pueda obtener la audiopágina VoiceXML con los contenidos similares pero adaptados al entorno auditivo.

7. Los negocios serán también otro elemento que formará parte importante de los portales de voz, ya que permitirá que los usuarios puedan tener acceso a adquirir no solo servicios sino productos. Claro está que para ello se necesita contar con un alto grado de seguridad en las transacciones. Sin embargo, como se pudo observar en el caso de algunos países desarrollados, ya se cuenta con algunos medios para proveer seguridad en las transacciones, por lo que el futuro del comercio mediante este canal se encuentra a un paso.

RECOMENDACIONES

1. Un portal de voz puede proveer gran beneficio no solo a las empresas sino a organismos gubernamentales de ámbito nacional, regional o local, ya que puede difundir información entre los ciudadanos y permitir transacciones telefónicas.
2. Este tipo de servicio es recomendable para todas aquellas entidades que deseen proporcionar un servicio accesible a la mayoría de la población, ya que un portal de voz representa un paso importante hacia la creación de una “administración virtual” al servicio de todos los ciudadanos.
3. Las empresas que decidan poner en marcha un portal de voz deben estar conscientes de que se enfrentarán a un creciente número de clientes que demandarán amplia gama de servicios, seguridad, fiabilidad y novedades que mantengan su interés.
4. A las empresas que ya cuentan con cierta infraestructura (por ejemplo bases de datos y servidores) se les recomienda utilizarla como apoyo para los sistemas de reconocimiento de voz, lo que les permitirá un ahorro económico considerable y un mayor rendimiento de las inversiones de Software realizadas previamente.

5. Si el deseo de las empresas es automatizarse para proporcionar un servicio las 24 horas del día los 365 días del año, la recomendación es hacerlo mediante un portal de voz, ya que el costo por minuto de un sistema de voz es significativamente más bajo que el de uno que cuente con operadores. Así mismo, esta automatización les permitirá reducir el número de operadores, reduciendo la planilla, tener menores costos de contratación y disponibilidad de espacio.
6. Este nuevo canal es recomendable para la captación de nuevos clientes como, por ejemplo, usuarios de viajes y turismo quienes necesitan soluciones en todo lugar y momento. Y como se sabe, en todas partes es fácil encontrar teléfonos.
7. El implementar las transacciones por voz también es una opción recomendable principalmente para acceder a nuevos sectores del mercado, proporcionando a clientes potenciales canales de acceso instantáneo a los servicios en cualquier lugar.
8. A través de los portales de voz también se puede fomentar la “audiopromoción”, mediante una cuidadosa selección de locutores con tonos de voz adecuados, de contenidos y de diálogos eficazmente diseñados, permitiendo de esta manera el desarrollo de una eficiente arma comercial cuyo efecto no solo se verá en la reducción de costos sino en la apertura de nuevos canales de ingreso.

9. En el caso particular de Guatemala sería necesario hacerles ver a las personas las grandes ventajas que tendrían al utilizar el servicio, así como fomentar en las empresas el desarrollo de estas tecnologías permitiendo de esta forma la reducción de los costos de inversión, pues al haber varias empresas que proporcionan el servicio, el valor del mismo estaría al alcance de la mayoría de la población. También sería importante involucrar la automatización de ciertos servicios como atención al cliente, realización de pagos, entre otras, con lo cual se podría prescindir de largas colas o de la necesidad de estar personalmente en una oficina o esperar largo rato a que el operador conteste la llamada.

REFERENCIAS

1. Ingeniería lingüística
http://www.hltcentral.org/usr_docs/Harness/harness-es.htm
2. Revista: Nueva economía numero 89
<http://www.el-mundo.es/nuevaeconomia/2001/NE089/>
3. Portales de Voz - SIA Sistemas
<http://www.siasistemas.com/sitio2/0208.htm>
4. Informática 2.0 Sitios que escuchan a sus visitantes
<http://old.clarin.com.ar/suplementos/informatica/2000/08/23/t-01101d.htm>
5. Temas a Fondo La Vanguardia Digital
http://www.lavanguardia.es/cgi-bin/noticialvd.pl?noticia=voz_actores&seccion=temas
6. Portal Eurociber
<http://www.eurociber.es/index.php?mostrar=xml4>
7. Ydilo servicio de asesoramiento
http://www.ydilo.com/esp/index_3c_2.html
8. Estos datos fueron proporcionados en el año 2003.

BIBLIOGRAFÍA

- Acceso Vocal a contenidos Internet
<http://www.tid.es/presencia/publicaciones/comsid/esp/20/4XX.PDF> (18/09/2003)
- Desarrolladores y Estándar VoiceXML <http://www.loquendocafe.com/index.asp>
(18/09/2003)
- Editores VoiceXML.mht (enlaces a editores y otras herramientas para ejemplos en el área de desarrollo de paginas VXML) <http://www.wmlclub.com>
(10/10/2003)
- El Estándar VoIP.
<http://www.comunicaciones.unitronics.es/tecnologia/voip.htm> (05/09/2003)
- El habla como medio de acceso a la Sociedad de la Información
<http://www.uclm.es/lamusa/paginas/monografico/Llisterri.htm>
(18/09/2003)
- Estado del arte en tecnologías de Voz.
<http://www.tid.es/presencia/publicaciones/comsid/esp/20/8XX.PDF> (05/09/2003)
- Ingeniería lingüística
http://www.hltcentral.org/usr_docs/Harness/harness-es.htm (18/09/2003)
- La próxima revolución será la voz <http://www.imim.es/quark/num21/021103.htm>
(18/09/2003)
- Lección 21 voicexml
<http://www.galileo.edu/wp/display/810/> (20/09/2003)
- Mapa de técnicas para desarrollo Web
<http://www.di.uniovi.es/~cernuda/survey.html> (20/08/2003)
- Página del portal de voz de Bellsouth en Guatemala
<http://www.bellsouthim.net.pe/voicep.asp> (25/09/2003)
- Portal de voz <http://www.heyanita.com> (20/09/2003)
- Portal de voz <http://www.tellme.com> (20/09/2003)
- Portal de voz para España y Latinoamérica <http://www.ydilo.com> (20/09/2003)

Portales de Voz: Internet en el teléfono

http://www.tid.es/presencia/publicaciones/comsid/esp/19/ART_2.PDF
(18/09/2003)

Servicio de Telecomunicaciones

<http://pegaso.ls.fi.upm.es/servicios/transparencias3/sld021.htm> (20/08/2003)

Voice Extensible Markup Language (VoiceXML)

<http://physionet.cps.unizar.es/~eduardo/docencia/tvoz/voicexml/VoiceXML20.pdf>
(18/09/2003)

VoiceXML Forum (pagina principal de las descripciones de dicho estándar)

<http://www.voicexml.org> (18/09/2003)

Voz IP: Presente y Futuro de las Comunicaciones de Voz

http://www.aui.es/biblio/libros/mi99/19voz_ip.htm (05/09/2003)

Voz sobre IP

http://www.ebosa.cl/pags/soluciones/redes_wan/cont2_voz_ip.html (05/09/2003)

Voz sobre IP Corporativa.

http://www.ebosa.cl/pags/soluciones/redes_wan/cont2_voz_ip.html (05/09/2003)

APÉNDICE

Herramientas que pueden ser utilizadas para desarrollar aplicaciones de voz

WebSphere Voice Toolkit (herramienta de IBM)

Es un software que permite crear aplicaciones e-business con tecnología de habla que pueden accederse por teléfono convencional o dispositivos móviles.

IBM WebSphere Voice Toolkit incluye el estándar abierto VoiceXML, lo que permite que los desarrolladores no tengan que contar con habilidades especiales de aplicación de voz, así mismo podrán aprovechar la infraestructura existente para proporcionar acceso de voz a información en sitios Web, bases de datos e intranets corporativas. Este software contiene componentes fundamentales, tales como editores gramaticales, diseñadores de pronunciación y módulos VoiceXML Resuable Dialog, los cuales aceleran el desarrollo de aplicaciones de voz y permiten a los programadores diseñar aplicaciones con tecnología de voz, incluso sin necesidad de estar familiarizados con los mínimos detalles de programación de voz, tales como la gramática y la fonética. IBM WebSphere Voice Toolkit es parte de la línea del producto WebSphere Voice Server de IBM que utiliza tecnologías estándar del mercado, como VoiceXML y Java, a fin de relacionar el reconocimiento del habla, la telefonía y la Web, para desarrollar e implementar los servicios con tecnología de voz. Está disponible para las pasarelas IBM DirectTalk, Voice-over-IP de Cisco y la plataforma de portal de voz Dialogic. Esta herramienta es la primera que ofrece soporte de voz integrada.

La familia WebSphere de IBM es un software o el middleware de infraestructura de Internet que permite a las empresas hospedar, ejecutar e integrar las aplicaciones de e-business.

Para poder realizar algunas prueba la beta de WebSphere Voice Toolkit esta disponible para ser descargada gratuitamente, a través del sitio Web alphaWorks de IBM en: www.ibm.com/alphaworks o en www.ibm.com/software/speech.

IBM WebSphere Voice Toolkit incluye:

- Un ambiente de desarrollo integrado (IDE) que se ejecuta en el PC y proporciona la capacidad de manejar los múltiples pasos requeridos para crear una aplicación de habla
- Un editor VoiceXML, que proporciona revisión de sintaxis y asistencia en el contenido
- Un editor gramatical, que permite la revisión de sintaxis y asistencia en el contenido
- Un diseñador de pronunciación que genera la pronunciación a partir de un texto escrito o la ortografía, así como la posibilidad de crear pronunciaciones manualmente
- Un grabador de audio que permite la creación de archivos de audio a partir de un texto escrito y proporciona los medios para ejecutar un archivo de audio previamente grabado
- Componentes de VoiceXML Reusable Dialog, un código VoiceXML previamente escrito, para utilizarlo como componente fundamental en funciones de aplicación, tales como nombres, direcciones, URLs e información de tarjetas de crédito

Requerimientos:

- Un sistema que soporte el ambiente operativo Windows 2000
- 50 MB, como mínimo, de espacio en disco (esto aún está en estudio)

- Hardware opcional para probar el código de aplicación recientemente desarrollado: micrófono de calidad, tarjeta de sonido y cornetas soportados por el ambiente de prueba
- Software opcional para probar el código de aplicación recientemente desarrollado: WebSphere Voice Server 1.5 SDK u otro ambiente de desarrollo que soporte el navegador estándar VoiceXML 1.0, JSGF y/o el formato de gramática BNF.

VoiceGenie Genie IDE 6.0 (September 2003)

VoiceGenie Genie IDE (ambiente de desarrollo integrado) es una herramienta enteramente ofrecida para el desarrollo de una colección que es ideal para desarrollar, manejar y desplegar aplicaciones de voz. La característica fundamental de esta herramienta es que permite utilizar todos los niveles rápidamente para desarrollar y desplegar aplicaciones VoiceXML de alta calidad para las empresas.

Incluye características que permiten editar archivos básicos, archivos y proyectos dirigidos, ayuda para la codificación, validaciones y plataforma de simulación.

Requerimientos del Sistema

La versión actual de VoiceGenie Genie IDE requiere Windows XP/Windows 2000 con Service Pack 1. Windows NT puede soportar esta herramienta, sin embargo aún no se puede garantizar su adecuado funcionamiento. Para instalar esta aplicación usted debe tener el instalador de Windows.

Características:

- Un editor de VoiceXML y editor de gramática XML, con ayuda en la codificación a través de etiquetas, atributos y valores propuestos.

- Una herramienta de manipulación de etiquetas que reduce importantes errores e incrementa la productividad en el desarrollo.
- Sintaxis resaltada y autoindentación.
- Construcción de proyectos dirigidos con capacidad de ayudas en proyectos organizados y facilidades de desarrollo.
- Soporte para el desarrollo local y remoto
- Herramientas de validación de VoiceXML y Gramática XML
- Plataforma de simulación que permite que sea posible correr archivos locales y remotos
- Versión 5.2 del Interpreter VoiceGenie's VoiceXML ofreciendo completo soporte para VoiceXML 2.0
- Flujo del diálogo para una representación grafica de la aplicación
- Habilidades de localizar, correr y depurar incluyendo soporte de puntos de interrupción, número de líneas para fácil rastreo usando la ayuda de depuración. Y también permite ir paso a paso para poder depurar mejor la aplicación
- Soporte de TTS SAPI

Documentación para VoiceGenie Genie IDE, así como la lista de bugs encontrados pueden ser consultados en <http://developer.voicegenie.com/IDEhelp.php>