

3

PORTAL DE VOZ DEL CIUDADANO

María José Lucía Mulas
Línea de Negocio CRM
Telefónica Soluciones

Jorge Navas Elorza
Coordinador de Area. Subdirección General de Atención al Ciudadano
y de Asistencia a Víctimas del Terrorismo/Secretaría General Técnica
Ministerio del Interior

En el ámbito de las relaciones de las Administraciones Públicas con el ciudadano, el acceso telefónico sigue siendo la forma más extendida de acceso a la información.

Los potenciales usuarios de los servicios de información de las Administraciones públicas son todos los ciudadanos, y el dispositivo de acceso más extendido entre todos los ciudadanos es el terminal telefónico. Ya sea un teléfono fijo, un teléfono móvil, o incluso una cabina telefónica pública.



Esta ponencia trata de las posibilidades de gestión de este canal de acceso, en particular de las posibilidades de automatización de la atención telefónica .

La experiencia concreta, a través de la cual vamos a ver el nivel de automatización alcanzable, es la experiencia de PORTAL DE VOZ DEL CIUDADANO DEL MINISTERIO DEL INTERIOR.

Primero analizaremos las posibilidades tecnológicas de la automatización de llamadas, el “estado del arte” en este campo, para, en segundo lugar comentar cómo se han aprovechado estas tecnologías en el Portal de Voz del Ciudadano.

1. PORTALES DE VOZ: POSIBILIDADES TECNOLÓGICAS

Hoy en día los sistemas de respuesta telefónica automática están perfectamente integrados en nuestra vida cotidiana. Estamos acostumbrados a interactuar con ellos y ya no nos sorprende ser atendidos por una máquina al llamar a cualquier servicio.

Hasta ahora la interacción se llevaba a cabo fundamentalmente con el teclado del teléfono: menús muy guiados y opciones de respuesta limitadas (las posibilidades de los 10 dígitos del teclado del teléfono). Ahora empieza a hacerse familiar un nuevo tipo de interacción, basado en la voz, que está abriendo posibilidades enormes de nuevos servicios. Vamos a comentar las últimas tecnologías en las que se basan estos nuevos servicios, denominados Portales de Voz por analogía con los portales de Internet.

Los elementos que componen una solución avanzada de Portal de Voz son los siguientes.

1.1. Equipos de Interfaz Telefónica

En primer lugar es necesario un equipo (normalmente un servidor Windows), o varios equipos, con unas tarjetas de telefonía que funcionan como un grupo de teléfonos - analógicos, digitales, o IP. Estos equipos cumplen la función básica de conexión a una red telefónica: pública o privada (a través de una centralita digital o IP) para recepción o emisión de llamadas.

La tendencia actual es concentrar en estos equipos las líneas telefónicas (un servidor puede albergar hasta 60/90 líneas telefónicas) y el software que ejecuta la secuencia lógica del diálogo, mientras que las funciones de reconocimiento y de generación de mensajes de voz a partir de texto se realizan, en función de las dimensiones del sistema, en el mismo equipo o en otros equipos adicionales conectados por red.

1.2. Software de Conversión Texto a Voz

La síntesis de voz a partir de textos variables es fundamental en los casos en que se desea dar una información dinámica y variable, o personalizada.

Actualmente existen conversores de alta calidad, que aunque todavía suelen tener algún defecto ya sea en el ritmo, prosodia, o calidad de la voz, y no llegan a confundirse del todo con la voz humana, ofrecen la información con una dicción y entonación absolutamente claras.

El software de conversión texto a voz, que antes residía en microprocesadores específicos, se ejecuta ahora en cualquier sistema operativo estándar (como Windows, o Unix). Atiende, en modo servidor, las peticiones de conversión de ficheros o textos, a las que responde construyendo el correspondiente mensaje de voz. La generación es prácticamente instantánea. Durante una llamada se pueden generar mensajes a partir de páginas de texto enteras sin que el llamante perciba ningún retraso en la generación de una respuesta.

Las empresas de conversión ofrecen la posibilidad de personalizar las voces sintetizadas en base a grabaciones de un locutor determinado. La voz sintetizada sonará así parecida a la voz real del locutor. Esta técnica permite combinar en un mismo sistema voz pregrabada y voz generada de gran parecido, y ofrecer la información fija y variable con voz homogénea.

Actualmente hay buenos conversores en español, inglés, francés, alemán, etc. Y catalán, valenciano, gallego, eusquera, etc.. Las empresas de conversión texto a voz suelen tener la mejor tecnología en su idioma de origen. En las distintas lenguas españolas casi siempre son empresas españolas las que dan la mejor calidad.

1.3. Software de Reconocimiento de Voz

Los avances en tecnologías de reconocimiento del habla son los que han permitido dar un giro radical en la forma de automatizar las llamadas. Los tradicionales diálogos basados en árboles jerárquicos de hasta 10 opciones (las 10 teclas de teléfono), han dado paso a unos diálogos que se quieren aproximar al lenguaje natural.

¿Hasta dónde llegan realmente las capacidades de las nuevas tecnologías, un poco ambiciosamente llamadas de reconocimiento de lenguaje natural?

- Permiten reconocer elementos de grandes listas (del orden de 10 000 entradas). Y por tanto hacer preguntas como ¿Desde qué localidad nos llama? ¿Con qué persona desea hablar? ¿Qué valor de bolsa desea consultar? ¿A qué restaurante desea ir?
- Permiten reconocer uno o varios elementos dentro de una frase, siempre que la respuesta pertenezca a un contexto restringido. Por ejemplo, en una aplicación bancaria, el diálogo podría contener preguntas del tipo “¿Qué operación desea realizar?” Y reconocer en respuestas como “Quiero conocer el saldo de mi cuenta de ahorro” los elementos clave “saldos” y “cuenta de ahorro”.

- Permiten reconocer elementos de listas generadas dinámicamente en función de los datos de la llamada.

Un ejemplo de diálogo posible sería:

Muy bien. Oficinas para tramitar su pasaporte. ¿Quiere conocer las oficinas en Madrid capital o en otras localidades de la provincia de Madrid?

En Madrid Capital

¿En qué barrio?

Chamartín

Puede acudir a la oficina situada en ...

¿Desea consultar otras oficinas?

No, Gracias

La limitación básica de estas tecnologías reside en que exigen una inversión de tiempo importante en el diseño y ajuste de los diálogos y de los contextos de reconocimiento. En efecto, el reconocimiento está estrechamente ligado al diálogo, a la forma de preguntar las cosas. Para cada pregunta hay que definir el posible conjunto de respuestas esperado (denominado gramática), y validarlo, una vez desarrollada la aplicación, en base a la escucha de cientos de respuestas reales de los llamantes. Es una fase en el desarrollo de los portales de voz que se ha denominado fase de afinado y que tiene como objetivo mejorar conjuntamente diálogo y gramáticas de reconocimiento.

El tiempo de afinado depende de los distintos criterios. Si el diálogo es muy dirigido, por ejemplo: “Diga uno a uno los dígitos de su código postal” o “Diga únicamente el nombre de la provincia desde donde está llamando”, el tiempo de afinado es menor que en el caso de un diálogo muy abierto, por ejemplo: “¿Qué desea?”.

Si los usuarios potenciales del sistema son usuarios avanzados, el tiempo de afinado se reduce. Los usuarios avanzados saben cómo hablar con una máquina, articulan bien, responden de forma breve, dejan tiempo para que la máquina responda, escuchan bien las preguntas para saber cómo responder... Si el sistema se dirige a un abanico amplio de usuarios distintos el tiempo de afinado será mayor. En estos casos es mejor optar por diálogos más guiados que delimiten bien los contextos de reconocimiento.

La evolución de estas tecnologías va en la línea de simplificar la definición de gramáticas y reducir el trabajo de afinado (enfoques del tipo Hot Word, Say Anything, etc), así como a mejorar los resultados en entornos ruidosos etc..

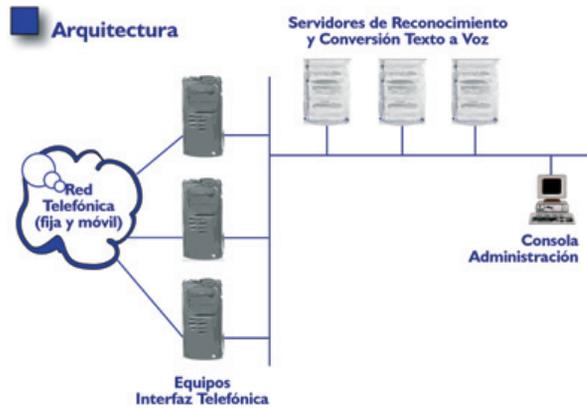
En relación con las tecnologías del habla hay que mencionar el software de Verificación del hablante. Se trata de un método biométrico basado en huellas de voz de forma análoga las huellas dactilares. El sistema verifica la identidad declarada por el llamante comparando su voz con las correspondientes muestras almacenadas en la base de datos. Este sistema se está empezando a utilizar en algunas aplicaciones del sector finanzas.

1.4. Arquitectura y Dimensionamiento

Las nuevas tecnologías de conversión texto a voz y de reconocimiento de voz trabajan en modo servidor. Este modo permite combinar conversores o reconocedores de distintas marcas en un mismo sistema, incluso en una misma llamada. Por ejemplo, en un sistema multilingüe se

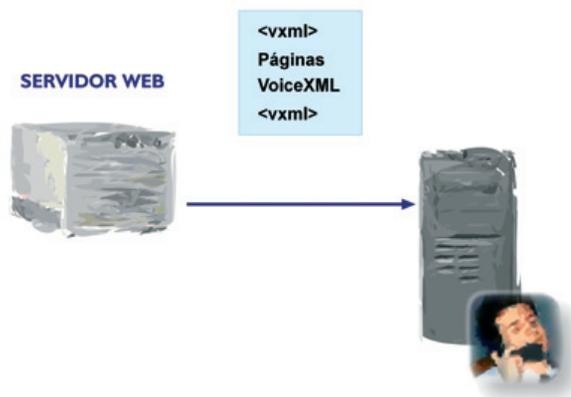
pueden dirigir las peticiones de conversión y reconocimiento a servidores distintos para cada idioma, y así atender las llamadas con la tecnología más adecuada.

La CPU necesaria para trabajar con estas tecnologías depende del número de líneas telefónicas que se quieran atender simultáneamente y de la cantidad de información que se desee generar o reconocer en cada llamada. El crecimiento de estos sistemas es fácil: en función de las dimensiones del sistema, el software de conversión y reconocimiento puede residir en la misma máquina que las tarjetas de telefonía, o, en máquinas adicionales, en cualquier punto de la red.



Las soluciones avanzadas de portales de voz ofrecen herramientas sencillas para administrar en red todos los elementos del sistema. La idea es que los administradores de los centros CRM dispongan de herramientas de fácil manejo para construir o modificar aplicaciones y parametrizar la ejecución de las distintas aplicaciones (por número de línea, por DNIS, en función de calendario, horarios, ...) en poco tiempo y sin necesidad de parar el servicio.

Estas soluciones también tienden a integrar herramientas para trabajar con Voice XML. VoiceXML es un lenguaje, basado en XML, para representar diálogos entre personas y sistemas automáticos, creado a partir de la idea de crear aplicaciones telefónicas con formato estándar que puedan ser cargadas desde distintos servidores de Internet. La aplicación escrita en VoiceXML reside en cualquier servidor WEB. El Portal de Voz en este caso ejecuta las aplicaciones de voz escritas en VoiceXML, que va importando de los servidores WEB, en función de las opciones seleccionadas por los llamantes.



1.5. Integración en los centros multicanal

Además de los elementos que acabamos de comentar no hay que olvidar que los Portales de Voz a través de la integración CTI y de la integración con los sistemas de backoffice se integran en los CRM multicanal potenciando las capacidades de enrutamiento de llamadas y gestión de la atención al cliente.

1.6. Costes

Para dar una idea de los costes de un sistema Portal de Voz, indicamos a continuación unos cuantos ejemplos basados en la plataforma avanzada MANAGERVOX de TELEFONICA SOLUCIONES. Los precios varían en función de la tecnología de conversión texto a voz o reconocimiento, y del tipo de aplicación.

HARDWARE Y LICENCIAS

4 LINEAS	25.000 € a 35.000 €
12 LINEAS	50.000 € a 75.000 €
30 LINEAS	90.000 € a 120.000 €

SERVICIOS APLICACIONES

CONSULTA DE PROGRAMAS DE OCIO Y CONSULTA DE ESTADO DE SOLICITUDES	40.000 €
PORTAL DE VOZ OFICINA DEL CIUDADANO	50.000 €
DIRECTORIO para 10 000 entradas	90.000 €

En función de la complejidad de la aplicación, a estos precios hay que añadir el coste de los servicios de desarrollo de la aplicación de voz y de integración con los sistemas de información.

El precio de los servicios varía mucho en función del tiempo de afinado y de la dificultad de integración con los sistemas de información.

2. EL CASO PRÁCTICO

Tras esta rápida revisión del estado actual de las tecnologías de Portales de Voz, vamos a ver cómo se han aplicado en el caso concreto del PORTAL DE VOZ DE LA OFICINA DE LA ATENCIÓN AL CIUDADANO Y DE ASISTENCIA A LAS VÍCTIMAS DEL TERRORISMO DEL MINISTERIO DEL INTERIOR.

2.1. Descripción del servicio

La Oficina de Atención al Ciudadano y de Asistencia a las Víctimas del Terrorismo es una Unidad adscrita a la Subdirección General de Atención al Ciudadano y Asistencia a las Víctimas del Terrorismo, dependiente de la Secretaría General Técnica del Ministerio del Interior, a la que compete la función de informar, dirigir y orientar al ciudadano respecto a los servicios y trámites prestados por el Ministerio del Interior, así como la de atención a las víctimas del terrorismo.

Este servicio se realiza fundamentalmente mediante un sistema de comunicación de tipo telefónico, abierto a todos los ciudadanos, y también con sistemas de información escrita, presencial y telemática, y a través de las publicaciones de todo tipo que tenga autorizado editar.

La Oficina aspira a dar servicio en condiciones de igualdad a todos los ciudadanos del territorio del Estado, por lo cual tiene como objetivo primordial la atención a través del teléfono, del servidor Web de Internet, y del correo postal y electrónico, como medios prácticamente gratuitos que les evitan el tener que desplazarse físicamente para ser atendidos.

2.2. Problemas y necesidades de la atención telefónica

En el año 2.000 la oficina de información al ciudadano no tenía ningún avance tecnológico a la hora de atender llamadas telefónicas. El único sistema implantado era un ACD (Distribuidor Automático de llamadas) que se encargaba de mantener encoladas las llamadas que no podían atender los operadores.

Los diez operadores se encargaban de cubrir turnos seis días por semana, y no podían atender el gran número de llamadas existentes, obteniendo un nivel de respuesta muy bajo, en torno al diez por ciento. Además, sólo se podían atender llamadas en horarios en los que estuviese abierta la oficina.

La información suministrada telefónicamente era muy apreciada por los usuarios, ya que era una información de calidad y así lo reflejaban las encuestas realizadas, pero no se podían atender tantas llamadas (en el año 2.000 se atendieron 220.000 consultas telefónicas) con tan poco limitados recursos técnicos.

A la vista de la situación, se decidió implantar un sistema automático de telefonía interactiva (IVR) con los siguientes objetivos:

- Elevar notablemente el porcentaje de las llamadas atendidas
- Proporcionar información ininterrumpida (24 horas x 7 días a la semana) automática de calidad; la meta clara era no quedarse en simples grabaciones estáticas, sino dar información personalizada.
- Utilizar sistemas avanzados en telefonía interactiva, tales como el reconocimiento de voz independiente del hablante y la síntesis de texto voz.

Se automatizó primero el número 902.150.002, dedicado en exclusiva para información de procesos selectivos, y posteriormente el número 900.150.000, número general de información administrativa del Ministerio.

2.3. IVR de Procesos Selectivos

El número 902.150.002 es el número oficial de información del Ministerio, aparece en las convocatorias en el B.O.E. y todos los opositores lo conocen.

Ofrece información sobre la oferta en vigor de empleo público del Ministerio del Interior, indicando en cada momento las órdenes y resoluciones aprobatorias de las diferentes convocatorias, fechas y lugares de los exámenes, documentación necesaria, consulta personalizada de resultados...

El sistema es un IVR tradicional, de 16 líneas analógicas, basado en respuestas DTMF, en el que se integró un sencillo interfaz para realizar un árbol de navegación vocal entre las distintas convocatorias en curso. En este interfaz se configuran, para cada convocatoria, los mensajes

que se han de emitir, los campos de la base de datos que hay que consultar, las indicaciones que hay que añadir en cada campo, etc..La aplicación telefónica genera los mensajes para cada convocatoria y candidato de acuerdo con los datos configurados.

Una vez la información de una oposición está cargada, las 16 líneas telefónicas asociadas a esta aplicación trabajan a pleno rendimiento; en algunos días se han llegado a atender hasta 11.000 llamadas personalizadas de manera totalmente automática.

2.4. Portal de Voz del Ciudadano

El segundo sistema que se ha implantado ha sido el de información administrativa general del Ministerio, el número 900.150.000, para información administrativa general del Ministerio.

Este número atiende llamadas que son más difícilmente sistematizables, ya que la información administrativa abarca todos los trámites y servicios prestados por las diferentes unidades del Ministerio del Interior: Documento Nacional de Identidad, Pasaportes, Extranjeros, Ayudas e Indemnizaciones, Tráfico, Asociaciones, Elecciones, Partidos Políticos, Armas, etc.

Para cubrir esta variedad se diseñó un sistema flexible con interfaces sencillas para que sus administradores pudieran incorporar paulatinamente menús diversos como:

- Emisión de información basada en mensajes variables, pregrabados, de texto o contenidos en ficheros de texto.
- Información de oficinas basadas en localización por provincia/localidades proximidad a códigos postales, configurable para cualquier tipo de oficinas.
- Información personalizada sobre cualquier tipo de trámites de forma similar a como se venía haciendo para oposiciones.
- Integración con buzones de voz y gestor de correo electrónico.

Ejemplos de automatizaciones posibles a partir de estas funcionalidades serían:

- Solicitar las oficinas para determinado trámite próximas a un código postal, o barrio.
- Pedir estado de una tramitación: el llamante daría su número de trámite y fecha de petición, o cualquier otro tipo de dato, y el sistema le ofrecería el estado de la tramitación y los siguientes pasos a dar.
- Pedir cita para previa para una gestión. El sistema solicita al llamante que deje un mensaje de voz con sus datos. El mensaje de voz grabado es enviado a una dirección de mail del gestor/gestores correspondiente.

Se decidió comenzar sistematizando información sobre dos de los trámites más solicitados: la información sobre el Documento Nacional de Identidad y la información sobre el Pasaporte.

Dada la diversidad de potenciales usuarios, y el hecho de que muchos probablemente no fueran usuarios avanzados, se optó por un diálogo bastante dirigido, utilizando intensivamente el reconocimiento de voz de la Universidad Politécnica de Cataluña, con opción final por DTMF cuando el llamante no consiguiera hacerse entender por el sistema con la voz.

Además se diseñaron las gramáticas de reconocimiento con criterios de sencillez, con la idea de que tanto los diálogos como las gramáticas fueran fácilmente modificables, y sirvieran de modelo sencillo para nuevos menús, ya que el sistema debía ser fundamentalmente dinámico.

El sistema se desarrolló con la **plataforma avanzada** de Portales de Voz de Telefónica Soluciones, **MANAGERVOX**.

Se dimensionó el sistema de forma que un mismo servidor con doble procesador albergara todos los elementos del sistema:

INTERFAZ TELEFONICA	12 LINEAS ANALÓGICAS
CONVERSIÓN TEXTO A VOZ	12 LICENCIAS CTV INTENSIVA ALTA CALIDAD DE SCANSOFT
RECONOCIMIENTO DE VOZ	12 RECURSOS RDV BUENACALIDAD (UNIVERSIDAD POLITÉCNICA DE CATALUÑA)

- El tiempo de desarrollo del sistema fue de 4 meses
- El coste total del sistema fue de unos 100.000 € (50.000 € Hardware y Licencias Software, 50.000 € Servicios)

2.5. Descripción de los beneficios obtenidos en funcionamiento normal

Los beneficios obtenidos son los siguientes:

Horario de cobertura del servicio: Al implementar este sistema, se ha pasado del antiguo horario de atención (días laborables de Lunes a Viernes de 9 a 17:30 y Sábados de 9 a 14:00), a disponer de un servicio ininterrumpido, 24 horas siete días a la semana, todos los días del año.

Tiempo de respuesta del servicio: El tiempo de respuesta del servicio ha mejorado notablemente, ya que anteriormente, el tiempo medio de cola en el ACD era de dos minutos, y la mayor parte de las llamadas entraban en saturación y no llegaban a entrar. Actualmente, excepto en picos puntuales, todas las llamadas se descuelgan y se les atiende de forma automática inmediatamente y, sólo en caso de pedir un trámite no automatizado, se transfieren a la cola de informadores del ACD.

Facilidad de contacto: El sistema implementado es muy amigable y fácil de usar, se han analizado cientos de conversaciones y más del 70% de los usuarios navegan por la información telefónica hasta localizar la información que buscaban sin abandonar el servicio.

2.6. Descripción de los beneficios obtenidos en situaciones de emergencia.

Capacidades del sistema

En situaciones de emergencia el sistema permite ofrecer de forma inmediata información urgente adecuada a las circunstancias.

Así ocurrió a raíz de los atentados del 11 de Marzo. El Portal de Voz permitió cambiar dinámicamente el servicio, sin tener que reiniciar el sistema en ningún momento, anulando la aplicación general de información sobre DNI, Pasaportes, etc e ir incorporando:

- 1) Locuciones de bienvenida con conversión de texto a voz con la información crítica de última hora: ayudas de alojamiento, vuelos de repatriación, ubicación de fallecidos, etc. El día 11 se cambiaron un mínimo de 10 veces, según llegaba nueva información.
- 2) Desvíos de llamadas según las nuevas temáticas a extensiones y grupos de ACD.
- 3) Grabación de mensajes de las personas que llamaban en horario nocturno o en saturación de los operadores del ACD. El sistema solicitaba que dejaran un mensaje con sus peticiones. Todos los mensajes grabados fueron escuchados y atendidos, lo que permitió no perder ninguna llamada.

Un día normal se reciben unas 1500 llamadas, el día 11 se recibieron casi siete veces más, en total 10.079 . Entre el 11 y el 14 de marzo sólo se atendieron llamadas del atentado. Hasta las 6 de la tarde del 11-M, el 90% de llamadas eran de información de familiares, y un 10% de ofrecimientos de ayudas. Los días siguientes fueron variando, llamaban familiares y víctimas con necesidades sociales (alojamiento, repatriación), con lo que se habilitaron menús especiales para cubrir todas las posibles consultas. En cuanto se hizo pública la regularización de inmigrantes subieron mucho las llamadas de inmigrantes. Pasado un mes del atentado casi todas las llamadas ya eran para información sobre el cobro de indemnizaciones por daños corporales.

El sistema nos permitió dar un mejor servicio a los ciudadanos en unas circunstancias muy especiales.

