

Servicios Web Ubicuos Activados por Voz

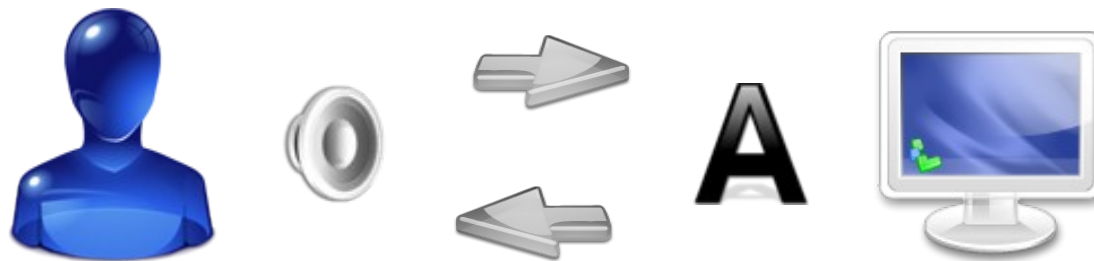
Parte I. Activación por voz

Ángel Gómez García
Dpt. Teoría de la Señal, Telemática y Comunicaciones
Universidad de Granada

Contenidos

- Interacción Oral Hombre-Máquina.
- Reconocimiento Remoto y VoiceXML.
- Reconocimiento Distribuido.
- Arquitectura multimodal Plus V.

Interacción oral hombre-máquina



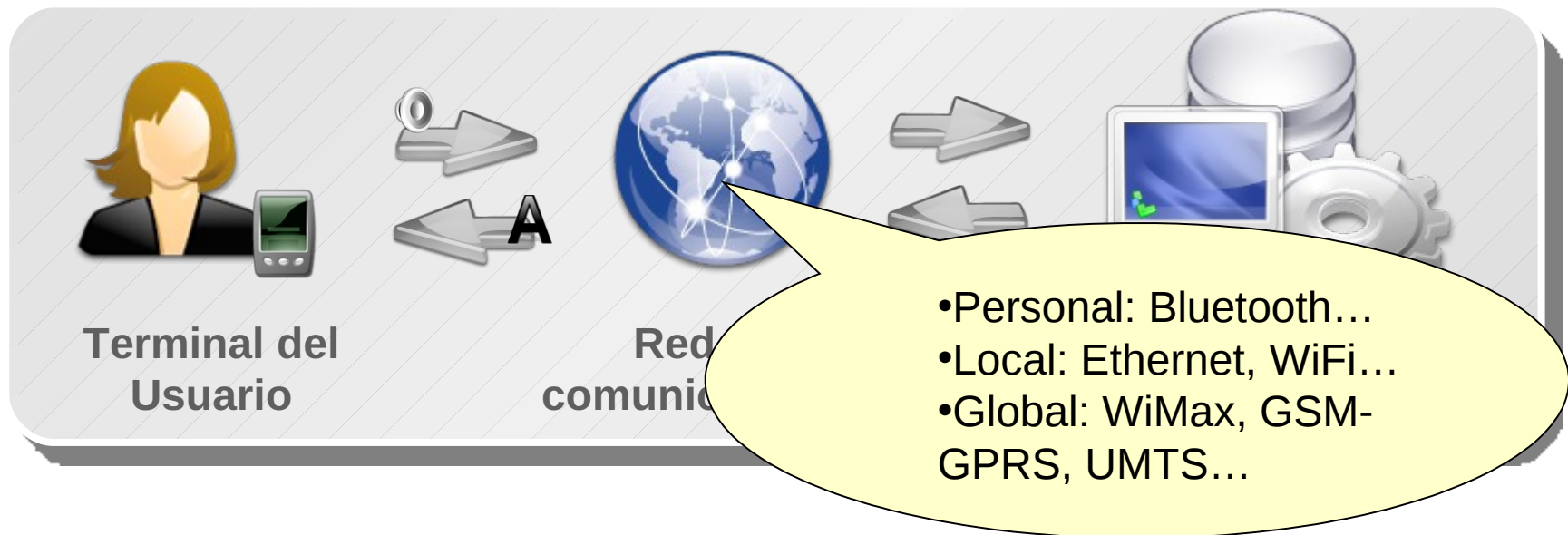
- Síntesis de voz: conversión de texto a voz.
- Reconocimiento de voz: conversión de voz a texto (o acción).

Aproximación tradicional al reconocimiento de voz

- Se realiza localmente, en el propio dispositivo.
- El usuario tiene el control absoluto.
- Depende de la potencia del equipo.
- El proveedor de servicios está limitado:
 - Realimentación positiva: servicios novedosos y útiles fuerzan a los usuarios a mejorar su reconocedor.
 - Realimentación negativa: ajuste a la baja para alcanzar al mayor número de usuarios.

Reconocimiento Remoto

- Se realiza fuera del dispositivo.



- Necesidad de una red comunicación.

Ventajas del reconocimiento remoto

- El usuario ya no tiene el control
 - ➔ Actualizaciones y mejoras automáticas.
- Mayor eficiencia y potencia computacional
 - ➔ Últimas tecnologías disponibles.
- Reconocimiento ubicuo
 - ➔ Cualquier dispositivo puede hacerlo
- Disponible actualmente
 - ➔ Servidores basados en VoiceXML

VoiceXML. Origen

- Mediados de los 90: AT&T explora la implementación de nuevos servicios Web.
 - ¿Por que no usar servicios activados por voz?
 - Ratón y pantalla se sustituyen por micrófono y altavoz.
 - Llamamos literalmente a la Web.
- Este modelo cliente-servidor fuerza la ruptura de los sistemas IVR (Interactive Voice Response) monolíticos en:
 - Interfaz de aplicación y
 - Lógica interna del servidor, integración sistema, bases de datos, etc.

VoiceXML. Origen

- AT&T's PLM
- HP's TalkML
- IBM's SpeechML
- Motorola's VoxML

HTML

HTTP/TCP/IP

Voice XML

- Estándar Abierto
- Partes intercambiables

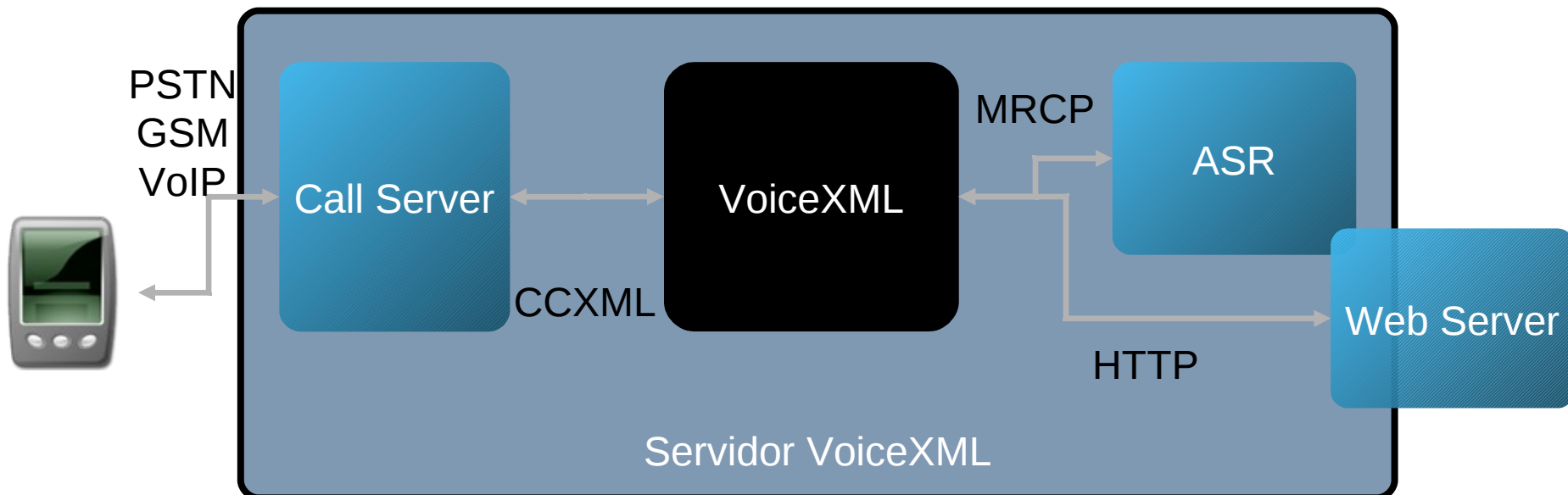


Secuencias
De Audio

HTTP/TCP/IP

VoiceXML

- VoiceXML es un lenguaje XML (eXtensible Markup Language) para gestionar la interacción por voz de la misma forma que HTML gestiona la interacción tradicional Web.
- La denominación se suele extender a todo el servidor.



VoiceXML

- VoiceXML permite a los desarrolladores definir-controlar el diálogo entre el usuario y la aplicación.
- Permite un amplio rango de acciones:
 - Síntesis de voz.
 - Reproducción de secuencias pregrabadas.
 - Reconocimiento de palabras o frases.
 - Identificación de tonos de marcado.
 - Grabación de voz.
 - Control del diálogo.
 - Control de la llamada.

Introducción a VoiceXML

- Hola mundo en VoiceXML

```
<?xml version="1.0" encoding="ISO-8859-1"?>  
<vxml version="2.0" lang="es" >  
  <form>  
    <block>  
      <prompt bargein="false">Hola mundo  
      <audio src="http://www.microsoft.com/sounds/ada.wav" />  
    </prompt>  
  </block>  
</form>  
</vxml >
```

Documento XML

VoiceXML 2.0 en Español

Organización en formularios

Síntesis del texto

Sin interrupciones

Reproducción de un sonido

Introducción a VoiceXML

- Una interacción sencilla:

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<vxml version="2.0" lang="es">
<form>
  <field name="servicio">
    <prompt>¿Que servicio desea buscar?</prompt>
    <option>Hotel</option>
    <option>Restaurante</option>
    <option>Gasolinera</option>
    <option>Hospital</option>
  </field>
  <block>
    <submit next="http://www.yip.com/call" servicename="servicio"/>
  </block>
</form>
</vxml>

```

Campo del formulario

Pregunta

Opciones dadas

Llamada al servidor con los datos

Los bloques presuponen la no interacción del usuario

Introducción a VoiceXML

- Servicio de páginas amarillas:

```
...  
<form name="yellowpages" >  
  <field name="servicio" >  
    <prompt>¿Que servicio desea buscar? </prompt >  
    <grammar src="servigram.xml #servicio" type="application/grammar+xml" / >  
  </field >  
  <field name="ciudad" >  
    <prompt>¿En que ciudad desea encontrarlo? </prompt >  
    <grammar src="ciudades.xml #ciudad" type="application/grammar+xml" / >  
  </field >  
</block >  
  <submit next="http://www.yo.com/call" servicename="servicio" city="ciudad" / >  
</block >  
</form >  
...
```

Podemos identificar cada formulario

Debemos rellenar dos campos para hacer la llamada

Uso de gramáticas (definidas mediante SRGS)

Introducción a VoiceXML

- Control de errores y ayuda al usuario

```
<prompt count="1" >
```

```
¿En que ciudad desea encontrarlo?
```

```
</prompt >
```

Contador automático de repetición

```
<prompt count="2" >
```

```
¿En que ciudad quiere encontrar un <value expr="servicio" / > ?
```

```
</prompt >
```

Respuesta por defecto

```
<nomatch>
```

```
<prompt >No entendí lo que dijo, ¿podría repetirlo?. </prompt >
```

```
<errorprompt / >
```

```
</nomatch >
```

```
<prompt >
```

```
¿En que ciudad?
```

```
<help>
```

```
Di ga el nombre de una ciudad. Buscaremos el servicio que desea en ella
```

```
</help >
```

```
</prompt >
```

Ayuda para la pregunta

Otras funcionalidades de VoiceXML

- Subrutinas.
- Operadores.
- Control de flujo.
- Control de errores.
- Tiempos de respuesta.
- Formularios de iniciativa mixta.

Desventajas de esta aproximación:

- Se envía más información de la necesaria.
 - Telefonía fija: 64 kbps, GSM-AMR: 12.2 kbps, G.729: 11.8 kbps.
- Se imponen más restricciones en latencia de las necesarias.
 - Hasta 500 ms son aceptables en reconocimiento.
- Efectos de codificación y compresión.
 - Afectan negativamente al reconocimiento.
- Aparición de ruido del canal:
 - Ruido transmisión, cortes, pérdidas de paquetes.
 - Intensifican los efectos de codificación-compresión.

Arquitectura Alternativa



- Realizar la transmisión una vez extraídos los parámetros que caracterizan el mensaje oral.

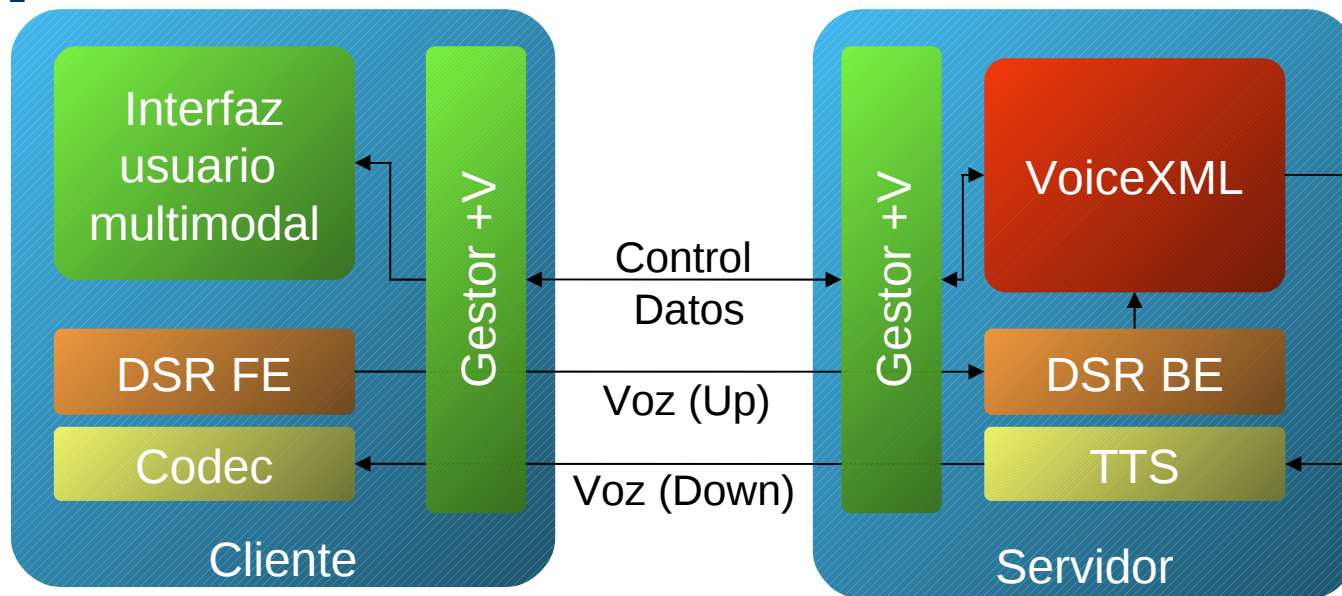
Reconocimiento Distribuido

- El proceso de reconocimiento se distribuye entre cliente y servidor.
 - Se elimina redundancia de la señal (4.8 kbps).
 - Reducción de ruido.
 - Protección frente a errores del canal.
- Estándares disponibles en ETSI:
 - Front-end estándar (FE), ETSI ES 201 108
 - Advanced front-end (AFE), ETSI ES 202 050
 - Extended front-end (XFE), ETSI ES 202 211
 - Extended advanced front-end (XAFE), ETSI ES 202 212
- Requiere cambios en el terminal para extraer las características de voz.

Arquitectura Plus V

- Propuesta por Motorola (promotores DSR).
- Objetivos:
 - Integración DSR en el esquema VoiceXML.
 - Permitir multimodalidad.
 - Minimizar los cambios.
- Independiente de la interfaz de usuario.
- Mueve el peso de la interacción al dispositivo del cliente.

Arquitectura Plus V



- El interprete VoiceXML pasa a gestionar subdiálogos.
- La capa de gestión +V se encarga de sincronizar el interfaz con la modalidad oral.

Enlaces útiles

- VoiceXML
 - <http://www.w3.org/TR/voicexml20/>
 - <http://www.w3.org/Voice/Guide/>
 - <http://cafe.bevocal.com/index.html>
 - <http://www.voxeo.com/prophecy/>
- Reconocimiento Distribuido
 - <http://www.etsi.org/WebSite/Technologies/DistributedSpeechRecognition>