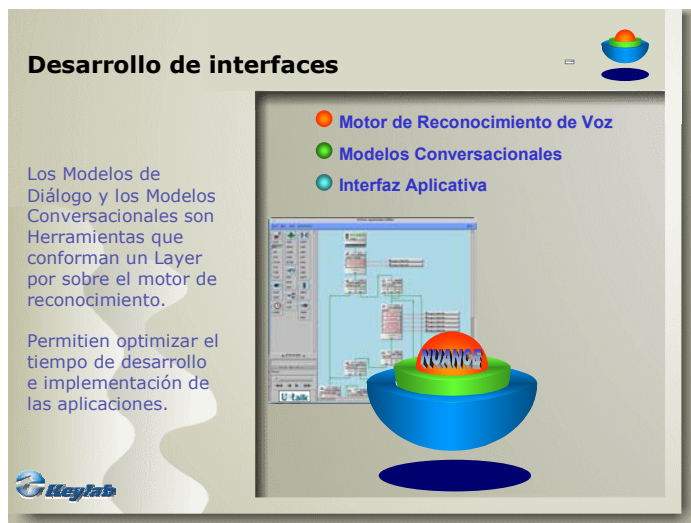


# Hacia una comunicación inteligente

Los sistemas de reconocimiento de voz permiten crear interfaces que reproducen los modos de interacción de los seres humanos. Si el armado del mensaje de bienvenida se diseña para que el usuario sepa desde un comienzo que se está comunicando con un sistema informático, pueden lograrse interacciones cómodas y efectivas.

Actualmente la tecnología nos permite interactuar con sistemas capaces de reconocer nuestra voz. Hoy es posible completar eficaz y satisfactoriamente transacciones a través del teléfono, desde las más simples hasta las más complejas, utilizando la vía de comunicación más natural y sencilla para la mayoría de los seres humanos: el habla.

La clave para que estas transacciones de voz sean exitosas depende de tener un 'engine' con la calidad, potencia y flexibilidad necesaria; contar con modelos conversacionales y modelos de diálogo especialmente diseñados para los intercambios verbales entre las personas y los sistemas de reconocimiento de voz; y, por último, de la adecuación de estos modelos a un contexto o población determinada.



Incorporar un sistema de reconocimiento de voz en una empresa puede significar **hacer más eficientes las comunicaciones, reducir costos operativos, proyectar una imagen de avanzada, estar a la vanguardia de las nuevas tecnologías e incrementar el capital de recursos humanos hacia otras áreas**, entre otros beneficios.

Sin embargo, tener la mejor tecnología es una condición necesaria pero no suficiente a la hora de generar una interfaz sensible a las necesidades de las personas que se comunican con una computadora, por medio de la voz, a través del teléfono.

## ¿Qué sucede cuando una persona ‘habla’ con un sistema de reconocimiento de voz?

Ante todo, como cualquier tecnología nueva, requiere de un período de adaptación de los usuarios, especialmente tratándose de aplicaciones de uso masivo.

El primer contacto que el usuario tiene con el sistema es a través del *prompt de bienvenida*. (Los prompts son aquellos mensajes del sistema dirigidos al usuario).

Si el prompt de bienvenida utiliza voz humana **debe aclarar explícitamente que es un sistema automatizado**; de no ser así, los usuarios razonablemente podrían confundirlo con un operador humano, o bien pensar que el sistema será capaz de comprender cualquier frase, aún fuera de contexto, al igual que una persona. Estas situaciones resultan inconvenientes ya que los sistemas actuales contemplan una cantidad finita de frases como posibles respuestas para cada pregunta<sup>1</sup>.

Un estudio hecho en A.T&T demostró que si los usuarios creen que están hablando con un operador humano sus pedidos son generalmente más largos y complicados. Cuando saben que se trata de un sistema, los usuarios **acotan la cantidad de palabras y se ajustan a decir específicamente aquello que el sistema les requiere** para completar la transacción.

**Figure 3. Average number of words callers used to state their requests (as a function of the initial greeting).**

Initial Greeting	Avg. No. of Words in Request
AT&T. How may I help you?	12.99
AT&T Automated Customer Service. How may I help you?	12.43
AT&T Automated Customer Service. This system listens to your speech and sends your call to the appropriate operator. How may I help you?	10.52
AT&T Automated Customer Service. This system listens to your speech and sends your call to the appropriate operator. How may I help you? (text-to-speech)	8.47

Los sistemas actuales de reconocimiento de voz permiten la utilización del *diálogo natural*, pero con una cantidad de palabras limitada por cada punto de reconocimiento.

Una vez que el usuario sabe que está comunicándose con un sistema informático, el ajuste se da naturalmente con el fin de obtener rápida y eficazmente el dato, acción, transacción o información requerida.

<sup>1</sup> En la actualidad hay tecnologías en desarrollo que permiten el reconocimiento de ‘gramáticas abiertas’ las que complementan el conjunto de frases reconocibles. Ejemplo: ‘Say anything technologies’ de Nuance communications.

Cuando no hay inconvenientes en la interacción y el sistema reconoce todo lo que el usuario dice, el resultado es la **satisfacción inmediata del usuario** y un impacto sumamente positivo.

En los casos en que el sistema no reconoce inmediatamente las palabras del usuario, entramos en un área denominada 'error handling / smart recovery' (manejo de errores / recuperación inteligente con respuestas anticipatorias respecto de las conductas más frecuentes y esperadas de los usuarios).

Estas herramientas, entre otras, se usan a través de prompts con características particulares. Su fin es evitar uno de los efectos más comunes en las personas cuando la 'máquina' con la que intentan operar no les responde como quisieran: **la frustración**.

A través de un buen manejo de errores, prompts cada vez más inductivos, y el tono de voz adecuado, un sistema puede acompañar al usuario a llegar a destino evitando o disminuyendo los niveles de frustración.

El desafío de los diseñadores de los modelos conversacionales es, por un lado, lograr que el usuario diga lo que el sistema necesita en el momento justo; y por el otro, generar un impacto positivo y agradar lo suficiente como para estimular el deseo y la voluntad de utilizar el sistema con frecuencia.

De esta manera estaríamos entrando en el terreno de la comunicación, donde, a través del diálogo natural, se realiza un **intercambio influyente de palabras**.

### **¿Qué se entiende por Diálogo natural?**

El *diálogo natural* es lo que permite una interacción sencilla, ágil, amigable y confortable entre los usuarios y los sistemas de reconocimiento de voz.

Aunque estructurado previamente, el sistema permite hablar naturalmente a través de diversos modos de preguntar y responder al usuario.

Por ejemplo, en una aplicación de derivación activada por voz cuando el sistema pregunta:

*S- ¿Con quién quiere hablar?*

El usuario puede responder de varias maneras y todas ellas serán reconocidas. Por ejemplo:

*U- Quiero hablar con Juan Pérez.*

*U- Necesito comunicarme con Juan Pérez.*

*U- Señorita, deme a Juan Pérez, por favor.*

El usuario puede responder como normalmente lo haría con una persona, y no debe esforzarse en pensar cuál es la respuesta que el sistema está esperando de él. De este modo, se genera además la sensación de que es él quien tiene el control del sistema y no al revés.

Según J. Larson en su artículo “Parts of Speech: Building Dialogues”, en los diálogos entre un sujeto humano y un sistema de computación es posible esperar tres diferentes situaciones:

1. Que el ser humano dirija la comunicación realizando la mayor parte de las preguntas.
2. Que el sistema dirija la conversación por medio de un diálogo inductivo.
3. Que la dirección del diálogo sea mixta donde cada parte tome la dirección por turnos.

El primer caso es el más dificultoso de prever en la elaboración de diálogos ya que es imposible anticipar todo lo que la gente puede decir en cada circunstancia.

La tercera situación, donde la dirección del diálogo es compartida por turnos, es la ideal y hacia donde apuntan tanto el diseño de los diálogos como la tecnología. Pero tomará un tiempo hasta que los sistemas sean capaces de interpretar señales básicas como tonos de voz, ritmos de la conversación, silencios, risas, entre otros, para así llegar a una comunicación más eficiente, eficaz y certera (arribar a un concepto denominado “Affective Computing”).

Es justamente la segunda situación la más utilizada en interacciones telefónicas. El usuario es interrogado en una secuencia ordenada por los requerimientos y debe responder con palabras o frases sencillas.

Según el autor, este tipo de diálogos tiene tres beneficios técnicos:

1. Induce al usuario a concentrarse en responder una pregunta específica. No deja lugar para que se adivine la información que el sistema necesite recibir.
2. Permite que el sistema de reconocimiento de voz interprete en cada punto del reconocimiento un grupo reducido de palabras en cada frase, en lugar de un vasto vocabulario.
3. Reemplaza el estilo de diálogos utilizados en los típicos sistemas DTMF -Touch Tone- (“Si quiere hacer tal operación presione 1, etc.”), que el usuario deja de usar por incómodo y poco amigable.

### ¿Cómo se logra un diálogo amigable, ágil y confiable?

Un diálogo con estas características se logra otorgándole un **perfil de personalidad** al sistema.

Kate Dobroth en su artículo *Beyond Natural: Adding Appeal to Speech Recognition Applications?*, asegura que **la gente tiende a antropomorfizar\* a los ordenadores y a responder socialmente a sus aplicaciones**. Por ejemplo, ante un silencio prolongado del sistema aparecen respuestas tales como: “Se enojó y ahora no me va a querer responder”; o, “¡¿Qué le pasa a esta máquina, se volvió loca?!”.

---

\* Antropomorfizar es otorgar cualidades humanas a algo que no lo es.

Esto daría cuenta de la necesidad de “humanizar” los sistemas de reconocimiento de voz. Es justamente a través de la calidad de la voz y del contenido de los prompts que el sistema muestra o transmite una determinada personalidad, la que puede variar acorde a las necesidades de la aplicación y el contexto.

Respecto de la voz, sabemos que ésta transmite muchos elementos más de los que se pueden tener en cuenta en un análisis superficial. Hay al menos 4 elementos clave que se deben considerar cuando se elige una voz para grabar los prompts de una aplicación:

- Género
- Calidad de la voz
- Ritmo / velocidad
- Entonación / Pronunciación

No olvidemos que también **a través de la voz una empresa puede transmitir a sus clientes la imagen que desea** (seguridad, confianza, agilidad, etc.).

Respecto al contenido de los prompts, se debe trabajar sobre aquellos detalles que hacen que el usuario perciba al sistema como un asistente competente. Entre otras cosas, que no se disculpe demasiado si comete errores sin consecuencias, que no hable muy lento ni resulte excesivamente explicativo, etc.

Cuando el flow del diálogo está preparado para enfrentar estos comportamientos con preguntas adecuadas, que induzcan al usuario a seguir adelante, con respuestas inteligentes y un buen manejo de errores, se obtiene un impacto sumamente positivo.

De esta manera, el usuario percibe la *personalidad* del sistema y eventualmente la imagen que la empresa desee transmitir a través de los modelos conversacionales. Resultado: confianza y obtención de un usuario leal y a largo plazo.

En síntesis....

- ✓ La nueva tecnología de reconocimiento de voz permite crear interfaces acordes a los modos de interacción de los seres humanos.
- ✓ Es muy importante que el armado del prompt de bienvenida se diseñe de tal modo que el usuario sepa desde un comienzo que se está comunicando con un sistema informático.
- ✓ Es a través del diálogo natural que se logra generar interacciones cómodas y efectivas entre los usuarios y los sistemas.
- ✓ Darle *personalidad* a un sistema de reconocimiento de voz ayuda a disminuir el impacto de la frustración cuando el mismo no reconoce correctamente en varias oportunidades seguidas.
- ✓ Es muy importante que ante reacciones comunes, como los silencios deliberados de algunos usuarios, los modelos de diálogos y los modelos conversacionales contemplen un buen manejo de errores y una recuperación inteligente con respuestas anticipatorias.
- ✓ Los sistemas actuales pueden ocupar el lugar de un asistente virtual agradable, ágil y confiable, si se utiliza la voz y los prompts adecuados.
- ✓ Hoy es posible automatizar transacciones de un modo que además de ser efectivo y certero, genere empatía en los usuarios.
- ✓ Hoy... podemos hacer que las computadoras 'hablen' con las personas a través de una comunicación más inteligente.