Understanding Voice User Interface Design

Table of Contents

Controlling a machine by simply talking to it was science fiction only a short time ago. Until recently, this area was considered to be artificial intelligence. However, with advances in technology, Voice User Interfaces (VUI) have become more commonplace, and people are taking advantage of the value that these hands-free, eyes-free interfaces provide in many situations.

Creating intuitive and usable VUIs involves coordinating a large number of research-based best practices, with the unique needs of each project. This white paper explores the following factors that influence VUI design:

- User
- Task
- Cognitive load
- Persona
- Natural language (prompting)
- Natural language (understanding)
- Effective error recovery.

User

A VUI should be designed according to user characteristics. Will there be a lot of new one-time users, e.g., callers to a company? Or will there be mainly frequent users, who interact with the system many times a day, as in the case of a messaging system?

For novice users, a system has to provide appropriate instructions and anticipate their questions well enough that they get comfortably to their goal. Prompts for ordering pizza might include "how many pizzas"; "what size" and "what toppings do you want".

A system that will have frequent expert users won't need as much instruction, can use more complex commands, and must get tasks done in a minimum number of steps. For example, an advanced user command might be: "I want to order one large pepperoni pizza." This command gets everything done in a single step.

Ideally, a VUI should provide enough instructions for novices, while not holding back experts who already know what they want.

Task

It is important to understand exactly what the goal is for a system (e.g. to place a phone call, buy a plane ticket, or update your work schedule) and to determine the most efficient and intuitive way to get it done. You need to ask questions like:

- How does the typical person do this task?
- Is there a way to do it in fewer steps?
- How do most people picture this task, i.e., their "mental model" of the task?

By carefully mapping out these factors, a VUI can work with the way people typically approach a problem, rather than working against it.

Cognitive Load

Unlike vision, where your eyes can easily scan a scene repeatedly, speech comes in as a single stream of sound. You must rely on your memory to know what happened a few seconds, or a few minutes, in the past. This means that VUIs must carefully manage how much pressure is put on the user's memory. Keeping this "cognitive load" to a reasonable level is crucial in making a VUI easy to use. Giving the user appropriate amounts of feedback, as well as splitting up a complex task into an intuitive sequence of smaller interactions – taking advantage of that "mental model" – can go a long way to reduce the cognitive load.

Persona

"Persona" refers to the personality that the system portrays by the choice of voice actor, the tone of the wording in the prompts, and way the system interacts with the user. These choices depend on the target user group and the nature of the company hosting the system. For example, a restaurant with a beach theme many want a loud "surfer dude" persona for a system taking reservations, but a stock brokerage would likely prefer a more quietly polite and professional persona. For many applications, a more neutral polite/professional persona is the best choice.

Natural Language (Prompting)

While most VUIs are not like a conversation with another person, by making use of certain aspects of conversation, the interaction can be much smoother and more natural for the user. Appropriate pacing and turn taking can keep

the transaction moving forward, as well as keep the user interested. For example, make it clear when the user should respond, and don't keep them waiting with huge chunks of text. Also, restricting or widening the vocabulary in the system prompts can encourage or discourage the user to do the same, e.g., always using exact terms versus allowing synonyms for the same concepts; open-ended versus closed-ended questions.

Natural Language (Understanding)

This is where most of the magic happens. Everything up to this point helps to make a system easier for the user to understand, and to help the user make appropriate responses. Anticipating all the different thing a user could say is the job of a well designed grammar. The grammar determines what the machine can 'understand', and then act upon. An extremely limited grammar may accept only the exact words "yes" and "no", while a more complex grammar might also correctly accept "yes please", "uuuuhh…yup", or "yeah, baby!" An even more complex grammar may also accept "I don't know", and then follow a different set of prompts to help out the user who needs more assistance.

Effective Error Recovery

Unfortunately, some errors or misunderstandings can occur, just as they do in regular conversation. By anticipating the most common errors, and giving the user effective and escalating instructions and feedback, the VUI can get back them back on track as smoothly as possible.

Examples

If a user remains silent, apparently not knowing what to do, after a short interval the system could give more detailed instructions:

System: "Please enter your PIN.

User: <silence>

System: "Please enter the four digit personal identification number that you chose earlier."

If a user says something that is not in the system grammars, leading to a recognition error, the system should let the user know what kind of input is required:

System: "Who do you want to speak to?"

User: "Bob."

System: "Sorry, I couldn't find that name. Please say the first and last name of the person you want to call."

User: "Bob Jones."

The idea is to give the user the appropriate level of instructions to keep them moving toward their goal, and in as natural a manner as possible. In many cases, the user would not be aware that an error occurred, and will just hear the next prompt in a logical sequence. If things are clearly not progressing, the VUI can even offer to transfer the user to a live agent before the user gets too frustrated.

Pronexus Speech and IVR Consulting

The technical side of speech recognition has matured to a point where correctly capturing spoken input is very reliable. Now a major challenge is creating intuitive and efficient VUI to properly harness this technology as a tool to get real work done in the real world.

Developing an effective VUI involves a number of activities that a typical software developer may not have the time or human-factors expertise to tackle.

This is where Pronexus VUI consulting services can help you succeed. We can design a new VUI from the ground up, edit and optimize an existing voice application, or even adapt an IVR (touch-tone) system to migrate it effectively to get all the advantages of a VUI.

Pronexus VUI consulting services include:

1. **VUI design**:

   - Task analysis/requirements gathering

   - Interpret requirements into sample dialogs for demonstration (user centered design approach)

   - Call flow and dialog design (typically as a Visio® and Word® document)

   - Produce a list of prompts (have them profession-ally recorded if desired)

   - Develop and tune grammars

2. **VUI review:**

   - Analyze an existing application's call flow and dialogs for inconsistencies or weaknesses

   - Usability assessment, to point out likely problems

- Produce a report of suggested changes to optimize VUI (typically as a Visio and Word document)

3. **Tuning services:**

   - Analyze an existing application's dialogs and grammars

   - Re-evaluate an existing call flow's success in achieving its goal

   - Optimize recognition parameters

Pronexus can provide individual services, or can produce an entire application.

Summary

Some would say that a great VUI is one that people don't remember using – it seamlessly got the job done. A good VUI can help overcome user aversion to technology because there are no new skills to learn -- people already understand how to make a phone call. Users can perform transactions quickly and easily, as though they had been speaking to a knowledgeable person. When care is taken designing a VUI, both the user and the organization implementing the automation win: the use has more control over conducting business at their own convenience, while the organization benefits from improved productivity and enhanced service delivery. With today's high recognition rates, VUIs have moved from the realm of science fiction, to providing reliable tools that are ready for use right now.