

While there are many important elements in a speech application, the voice user interface (VUI), which is the portion of the system that callers hear and speak to, has potentially the greatest impact on achieving a rapid, positive return on investment. Callers will only use speech applications that meet their needs and are easy to use. To optimize your VUI, start by asking yourself two important questions:

1. *What should I automate?*
2. *How do I make it the best it can be?*

Choosing the right application is crucial for ensuring success. Typically the best business benefits are derived by targeting fairly simple tasks in a constrained environment with high client traffic.

As for how to go about it - read on. Speech experts from 20 different companies have pooled their knowledge to bring you some of their favorite suggestions for developing VUIs. We encourage VUI designers everywhere to apply these suggestions. We encourage managers of voice application projects to support their VUI in applying these guidelines.

1. You can't design what you can't define

Most automated transactions are representations of real world transactions between a caller and a live agent. Examining the task from as many angles as possible will help the designer arrive at a sufficiently detailed task model. The model will include the set of logical steps that make up the caller's interactions with the system and will support the definition of requirements for the recognition vocabulary and the system's dialog flow.

In VUI design, as in any software project, requirements derive from two sources: the business goals and a clear understanding of customers and their needs. On the business side, VUI projects are truly multi-disciplinary – requiring participation from a number of stakeholders including marketing, business, IT and customer service. Examining the task and business goals from these diverse vantage points will help generate a complete set of requirements.

The flip side of the design equation is the caller and task analysis process. It is important to follow a defined analytical process that examines a wide range of caller characteristics and task functions. Careful consideration must include a detailed understanding of the callers, their environment, the various interactions involved in the task, and of course, the desired outcomes and behaviors of the VUI.

2. Use caller-centered design techniques.

At SpeechTEK 2004, a group of leading VUI designers attended the Voice User Interface (VUI) workshop directed by Dr. James A. Larson. Taking the lead for an article on the best practices in VUI, Dr. Larson collected and coordinated this team of VUI specialists to compile the Ten Guidelines for Designing a Successful Voice User Interface. Speech Technology Magazine would like to thank the authors for their contributions to this article and Dr. Larson for his efforts to coordinate this joint effort. Thanks go to:

Ted Applebaum
Bill Byrne
Michael Cohen
James Giangola
Juan E. Gilbert
Rebecca Nowlin Green
Thomas Hebner
Tom Houwing
Susan Hura
Sunil Issar
Lizanne Kaiser
Karen Kaushansky
Robby Kilgore
Jennifer Lai
James A. Larson
David Leppik
Stephen Mailey
Ed Margulies
Kristen McArtor
Michael McTear
Richard Sachs

In gathering requirements for design, make sure the caller is always in the driver's seat. This entails understanding and respecting the caller's expectations. The design also needs to take into account a variety of usage environments (e.g... mobile, inside/outside, or accessibility issues) so that the VUI will work for a wide range of callers. Make sure to drive transactions to conclusion without cluttering the experience with distractions. For example, rather than organizing a dialog based on how the company is organized, or how internal departments interact with each other – find out how callers think about things, and design the VUI to match their perception. This means avoiding jargon and anything else that might hinder the caller's agenda. (Research¹ shows that 5.1 percent of users have trouble with jargon and shop talk in speech dialogs.

3. Use the right technology, and use technology right.

Use the right technology: Choose the technology appropriate for the task to be completed. For example, if you must mix synthetic speech with recorded human speech, use synthetic speech for reading dynamic text, such as e-mail messages, weather reports, or sports scores, and pre-recorded human speech for the static system prompts. As another example, use a statistical language model, which functions like a grammar when mapping speech to text, but works from a statistical model rather than a set of fixed rules, when trying to route calls to a wide array of destinations

Use the technology right: It's equally important to understand how to configure the technologies you are using. For example, the smallest change in a parameter setting can have a noticeable effect on the performance of your VUI. Choose the appropriate dialog strategy². Remember that speech recognition, dialog interaction, and back-end application processing may occur on different servers. If there is narrow bandwidth between these servers, there may be latency which your dialog must accommodate.

4. Leverage the language instinct

The role of conversational language in VUIs is not to fool callers into believing they are talking to a real person. On the contrary, efforts to create this false impression are likely to have disastrous consequences for both recognition rates and caller satisfaction. Rather, the reason for using a conversational metaphor is to leverage a caller's natural "language instinct,"¹ a mental faculty that took hundreds of thousands of years to evolve. Designers should use this faculty for language use to their advantage in order to comfortably and comprehensibly usher callers through whatever the VUI task may be.² For example, opt for wordings that will be familiar to callers; avoid grammar practices that typify written literature rather than everyday spoken language; use cohesive devices such as pronouns, acknowledgments, transition words³; record prompts with familiar, natural intonation contours⁴ and enable callers to speak over and interrupt messages that they may not need to hear. (Research⁵ confirms that 12.8 percent of callers felt that the system seemed to not shut up or want to give the caller a turn.) VUI usability research confirms that the use of natural, contextually appropriate prosody—intonation, stress, rhythm—significantly improves callers' recollections of information as well as the likeability of the system⁶.

5. Invest in quality: establish success criteria and test against them

Testing the IVR makes sense for the same reason it makes sense to test your core products and services; VUIs represent a significant investment. Testing provides a low-risk preview of performance and reactions to your VUI. Discovering problems early in the project lifecycle saves time and money. System performance testing includes recognition performance (how accurate is the recognition for a variety of people in a variety of settings), stress testing (how

many callers can use the system at the same time), dialog traversal testing (testing many possible conversations), and host connectivity. Usability testing addresses the quality of the caller's experience with the system. In usability testing real callers interact with a realistic version of the VUI to complete typical tasks. This produces data on how successful callers are as well as their reactions to the system and their opinions of it.

To ensure the success of the final VUI, test early and often against established success criteria. This allows you to focus on the right goals during the various stages of testing: prototype Wizard of Oz testing, pre-deployment usability testing, and post-deployment development testing/tuning. Turn established business goals into testable criteria. Specify success metrics determined during the requirements for testing by real customers. Some examples of success metrics include call durations, single-call completion rates (how many callers complete their task on their first call), caller satisfaction (how many callers report being ?very satisfied?), and any change in brand loyalty (how likely are callers to remain your customers because your over-the-phone customer service is better than what they expected from your competition).

6. Branding in VUI is more than just a pretty voice.

Since the 1960s, sociolinguistic researchers have repeatedly demonstrated that humans infer a vast inventory of socio-economic, geographic, and personality traits from brief speech samples⁷. VUI callers are likewise bound to perceive such traits while interacting with the voice of your application—whether you plan for it or not⁸. The VUI provides an opportunity to leverage existing investments in branding and differentiation with respect to competitors' offerings. It is important to define the personality of that voice early in the project because it affects dialog structure, prompt wording and the tone of the recordings. At its most basic, thinking about personality makes it easier to avoid sounding like a robot. After all, how often do people ask you to "input" your phone number? It is important that the personality reflects both the company's brand image and the preferences of the callers. Without acceptance from the caller, the brand just won't work.

7. How you say it is as important as what you say

Even the best-designed VUIs will fail unless their carefully crafted conversational design is captured in the prompt recording session. Voice acting, directing and post-production are specialized skills that require specific technical and design expertise as well as artistic talent⁹. Choosing the right voice actor will ensure that the prompts are delivered with accuracy and the appropriate tone. Likewise, an experienced director who knows the application will make sure that all the subtleties of spoken language such as stress and intonation are brought to life in the VUI. For example, phrases that include dates, times, and dollar amounts are typically made up of several shorter prompts and are dynamically pieced together by the VUI. To get a date such as "January 1st, 2004" to sound natural, the individual prompt units must be recorded with just the right intonation so that they fit together seamlessly. Also, the way a prompt is spoken may affect how the user responds^{10, 11}. For example, "Are you a member or a provider?" may elicit a response of "yes" while "Are you a member? Or a provider?" may elicit either the words "yes," "no," "member" or "provider" as a response.

8. Don't block the exit

There is a trade off between preventing callers from reaching Customer Service Representatives (CSR) and caller satisfaction. In fact, allowing easy access to CSRs is much more advantageous than preventing such access. Data show that preventing the users from transferring to a CSR will adversely affect calls as well as opinions:

1. Callers who only want to speak to a CSR can be very determined and will do whatever it takes to reach a live CSR, including saying nothing (playing possum) in hope of an eventual transfer, coughing repeatedly, calling multiple times, or pressing zero repeatedly. These actions mean extra call minutes for both the system and the caller.
2. When they finally do reach a CSR, callers will spend extra time complaining about their automated experience. This is expensive time for both caller and CSR and callers will be less likely to want to use the automated system in the future.

Callers routinely complain about being forced to use automated systems. However, providing easy access to a CSR does not mean more transfers for a well-designed VUI. There are effective ways to make the transfer request a win-win situation:

1. Offer the caller a choice between continuing with self-service or waiting—with anticipated wait time—to reach a CSR.
2. Anticipate partial call automation. To save CSR time and minimize caller frustration, transfer what data has already been collected while routing the caller to the appropriate CSR for first-call resolution.
3. Send the caller back to the automated system after a CSR handles the caller's particular request.

From the caller's perspective, a company's call center is all one customer service experience, whether it means self service, assistance from call center staff, or a combination of the two.

9. Take care with error handling

Error handling can be particularly vulnerable to "conversation-breakers"—conversational break-downs that throw off the natural rhythm and upset the progress of the interaction, leaving the caller confused and at a loss. When an automated system repeatedly fails to hear or understand the caller's spoken request, frustration naturally focuses on the short-comings of the system. Callers will likely judge a system which consecutively repeats the same error message in the same way they judge people who simply repeat. This lack of "social memory" is perceived as both unhelpful and unintelligent.

To counter this negative perception, VUI designs may employ simple forms of "context awareness" which help move the conversation forward. By keeping track of a conversation (just as real people do) with phrases, such as "Sorry, but I still didn't understand", and offering ever more targeted help, systems can move the conversation forward in a way that is helpful to the caller. When simple strategies like these are overlooked, conversations lack context and feel awkward¹². Worse yet, callers hang up.

10. Establish a change process

All the elements of a VUI are highly interdependent, from low-level choices of prompt wording and call flow, to high-level choices of VUI features and system personality. As a result, even a seemingly small change—for example, a single word in a prompt—may have undesirable consequences. Avoid the temptation to "twiddle bits" just as you would avoid randomly changing statements in a complex computer program. If a change is needed, no matter how small, do it through a process that includes the VUI designer. The designer can examine the impact of any change on the interactions and consistency with other portions of the design.

Summary

Deploying a successful speech solution depends upon a defined process incorporating the above guidelines. In-depth requirements, gathering and analysis, careful design and extensive testing are essential in order to achieve an effective VUI with both high performance and usability scores. Voice user interface experts from 20 companies agree,

compromising on these guidelines will have undesirable effects not only where callers are concerned, but also on your VUI investment.

¹ Pinker, S. (1994) *The language instinct*. New York: William Morrow & Co.

² McTear, M (2004) *Spoken Dialogue Technology 2004*, Springer-Verlag.

³ (For a review of conversational design features, see Chapter 10 "Designing Prompts")
Cohen, M. H., Giangola, J. P., & Balogh, J. (2004). *Voice user interface design*. Boston: Addison-Wesley.

⁴ For natural prosody in prompting and concatenation, see Chapter 11 "Planning Prosody" in Cohen, Giangola & Balogh, 2004 *Voice user interface design*. Boston: Addison-Wesley.

⁵ Margulies, E. (Ed.). (2004) *Voice Response Usability Almanac 2004*. Las Vegas, NV: Sterling Audits.

⁶ Balogh, J. (2001). "Strategies for concatenating recordings in a voice user interface: What we can learn from prosody." *Extended Abstracts, CHI (Computer Human Interface) 2001*, pp 249-250.

⁷ For example: Labov, W. 1966. *The Social stratification of English in New York City*. Washington D.C.: Center for Applied Linguistics . Also: Giles H. & P. Powesland. 1975, *Speech style and social evaluation*. (European Monographs in Social Psychology). New York: Harcourt Brace. For a general review of ?speech evaluation? research, see Hudson, R. 1980. *Sociolinguistics*. London: Cambridge University Press.

⁸ For application of speech evaluation research to VUI design and a practical "Persona Design Checklist," see Ch. 6 "Creating Persona, by Design" in Cohen, Giangola & Balogh, 2004.

⁹ See Ch. 17 "Working with Voice Actors" in Cohen, Giangola & Balogh, 2004.

¹⁰ Byrne, Bill, "In the Studio: setting High Standards for Prerecorded Audio," *Speech Technology*, March/April 2004, http://www.speechtechmag.com/issues/9_2/cover/10195-1.html

¹¹ [For a review of intonation, stress, and concatenation strategies for VUIs, see Ch. 11 ?Planning Prosody? in Cohen, Giangola & Balogh, 2004.](#)

¹² [For a review of error handling strategies along these lines, see Chapter 5 \("High-level design elements"\) in Cohen, Giangola & Balogh, 2004; Chapter 10 \(Error recovery and prevention\) in Bruce Balentine & David P. Morgan, *How to Build a Speech Recognition Application: A Style Guide for Telephony Dialogues*, 2nd edition, 1999, San Ramon, CA: Enterprise Integration Group, and Chapter 8 in James A. Larson, *VoiceXML Introduction to Developing Speech Applications*, 2003, Prentice Hall](#)