

SoundHound Inc.

# Finding Your Brand's Voice: 6 Ways to Build a Better VUI



As the world moves toward more voice-enabled products and services, it's imperative for brands to have a voice-first strategy. Read this best practice guide for expert advice on building a better voice interface (VUI) for your brand.

# Expert Voices in This Guide



**Mihai Antonescu**  
Sr. Engineer  
Mercedes-Benz R&D



**Octavio Menocal**  
Sr. Voice Experience Engineer  
RAIN



**Ananya Sharan**  
Product Manager  
Pandora - Voice Mode



**Lauren Golembiewski**  
CEO & Co-founder  
Voxable



**Will Hall**  
Chief Creative Officer  
RAIN



**Heidi Culbertson**  
Founder & CEO  
Marvee



**Rachel Batish**  
VP of Product  
Audioburst



**Elissa Dailey**  
Senior Strategist  
RAIN



**Michael Mahmood**  
Director of Product Design  
SoundHound Inc.

# Table of Contents

## **Extending Your Brand Through Voice**

Pages 4-14

## **Infusing Personality and Humor**

Pages 15-25

## **Enabling Natural Interactions**

Pages 26-34

## **Prompting and Error Recovery**

Pages 35-42

## **Onboarding, Education and Discovery**

Pages 43-52

## **Personalizing User Experiences**

Pages 53-58

Chapter 01:

# Extending Your Brand Through Voice



# Defining Your Voice: Getting Vocal With Your Brand

The world is becoming more voice-enabled. By next year, 75% of all households in the United States will have some type of smart speaker supported by a voice user interface (VUI). By 2021, a whopping 94% of large companies expect to rely on voice automation for everything from customer service and sales, to helping employees managing schedules. The time is right to stop and think: how will your existing brand values and attributes translate to voice? In fact, what does your brand sound like? And how does a voice assistant interact with your customers in an impactful and meaningful way?

"Cutting through the clutter is hard for any brand, and doing so in a memorable way is even harder. With many voice assistants, we lose an important component of sense memory — sight — and rely mainly on hearing as a means to spark association, memory, and interaction," says Elissa Dailey, senior strategist at RAIN.

"Therefore, literal brand voice is even more important to define, differentiate, and master to ensure memorability and impact."

Voice technology opens up a new way for brands to build emotional connections

with their target audience on a deeper level. One way to do this effectively is through a well-defined brand voice that's consistent across channels, devices, and platforms. But where do you start to ensure that your voice strategy is on-brand, engaging, and genuinely helpful? And how do you teach your customers and clients to use it?

Chances are, you've already been through branding exercises to develop logos, colors, value, mission statements, and personas. In the world of voice technology, it's necessary to brand your voice as well. For example, companies like Pandora now offer a voice mode that allows users to collaborate with it to find music that fits their tastes and moods. It's a highly personal experience that depends on brand trust. From the jaunty "bloops" of a Skype call to hearing the familiar voice of a friend every time we summon Siri, audio branding is as complex and as important as visuals in today's world. Creating a visible identity worked well so far, but what happens as more and more brands start to talk?

**"With many voice assistants, we lose an important component of sense memory - sight - and rely mainly on hearing as a means to spark association, memory, and interaction."**



**Elissa Dailey**  
Senior Strategist  
RAIN  
@elidai

## Best practices:

### 1 Strengthen brand experiences with an easy-to-remember wake word

Audio branding is nothing new. From the familiar three tones of NBC-TV to the whisking-away sound of emailing in Microsoft Outlook, we're used to associating sounds with brands. Since virtual assistants and voice technology have come to the forefront, brands need to realize that this first encounter is the most important one of the interaction. When you say "Alexa," you know you're talking to Amazon, and you know that "Hey Siri" awakens Apple's voice-enabled software. But how can users get to your world?

Your voice roadmap is a critical extension of your brand. Some experts advise brands to partner with a known platform, like adding Alexa or Google to automobiles or appliances. But for brands that want to retain control over their brand and user experience, creating a custom wake word is the way to go.



Branded wake words enable brands to deepen user engagement, control the user experience and own their customer data.

For Motorola, the catchy “Hello Moto” phrase wakes its voice user interface on many of its smartphones. Mercedes-Benz owners are already summoning their in-car voice technology to do things like set a radio station, temperature, and seating position by saying, “Hey Mercedes...” And of course, a breezy “Hey, Pandora” sparks a personalized musical experience that’s a huge differentiator from its competitors. Connect the customized wake word with your brand itself, rather than inserting another brand wake word into the mix, and increase product name affinity.

## 2 Understand which voice truly personifies your brand

You’ve probably had experience creating user personas — the composite traits of a typical customer of your brand. You give them photographs, names, ages, and a back story. Have you ever tried going the other way and created a persona that represents your brand? If your brand were a person, who would it be? Would it be a serious and professional colleague, a friendly, chatty helper, or a funny, clever friend? What’s its name? What gender is your brand? What does the tone and pitch of its voice sound like? Shoe companies often hire sports stars to be the voice and face of their products because customers relate to characters most like themselves, or like their ideal selves.

***If your brand were a person, who would it be? Would it be a serious and professional colleague, a friendly, chatty helper, or a funny, clever friend?***

Austrian chatbot provider Onlim advises its customers to understand their consumer base and design the voice assistant to fit the customer personae. Consider the voice gender, age, race, and personality.

The smartest direction to take would be to craft your voice assistant based on a diverse set of customer data and explore options where users can select their preferred voice. This way, you can avoid potential biases and design an assistant that resonates with your audience and feels natural to them.

"For a lot of brands, especially big brands, you already have a brand experience your customers recognize. So it's really extending that same experience to another venue, through voice. It should always start off with, what's the business goal and who are you today and is this going to be multichannel or via a new channel," advises Heidi Culbertson, technologist, VUX designer and the founder & CEO of Marvee.

Pay attention to vocal characteristics when choosing a voice user interface to ensure that the character is right for your brand. Once you've done some market research to determine who your audience is, test some synthesized voice personas until you find one that matches both your audience profile and the image you want to project.

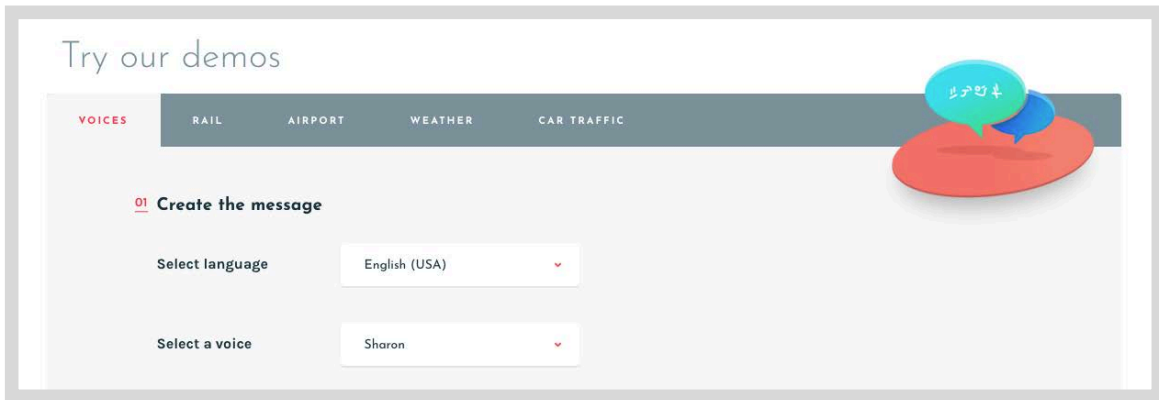


**Heidi Culbertson**  
Founder & CEO  
Marvee  
🐦 @hcultbertson

**"For a lot of brands, especially big brands, you already have a brand experience your customers recognize. So it's really extending that same experience to another venue, through voice."**

Today, synthetic voice creation requires audio recordings by professional voice actors in partnership with linguistic experts. But if you're looking to consistently create recordings of large volumes of personalized text, speech synthesis is a great choice and a quicker way to do this. Brands can also use a company like the Acapela Group to create a voice with only a few minutes of high-quality audio recordings and the associated text transcription of the audio samples (text-to-speech).





Text-to-speech voice synthesis [demo](#) on the Acapela website

### 3 Infuse the VUI with a character through tone and persona

"Conversations are social interactions," said Lauren Golembiewski, CEO and conversational and voice interface designer at Voxable. "Users often personify conversational interfaces and project human qualities onto them." People refer to both Siri and Alexa as "she," and talk to the interface conversationally. That quirk of humans—the same instinct that makes us see faces in random shapes—is key to the success of a voice assistant.

"We don't want to talk to a computer. We want to talk to someone who understands us, is empathetic toward us, and who talks like we do. To ensure the VUI has intuitive conversational interactions with users, designers must establish the VUI's voice, tone, and persona," Golembiewski continues. "Voice is the type and quality of the words the VUI uses. Tone is how those words modulate. And Persona is the character that embodies and informs the voice and tone. (These) are the brand foundations that drive what language designers and writers use in prompts, responses, and information the VUI generates."

**"You want it to be natural and conversational. Our vision was that the Pandora voice assistant is like that music expert friend that really knows you."**



**Ananya Sharan**  
Product Manager  
Pandora - Voice Mode  
@\_ananya\_

For Ananya Sharan, product manager of Pandora's Voice Mode, it's always been about helping people find what they want, more easily. "From the beginning, keeping our listeners in mind, we wanted the voice experience to be effortless. You want it to be natural, conversational. Our vision was that the voice assistant is like that music expert friend that really knows you."

"We wanted people to be able to ask for things like, "Play me something awesome," adds Sharan.

***Selecting a persona is just as important as selecting a brand ambassador and should go through just as stringent a vetting process.***

Your voice user interface needs to be the literal voice of your brand. It should speak with confidence and authority, keep responses brief, and provide the right amount of information. It should be friendly, familiar, and knowledgeable, and it should be able to expect what we might ask next. Dailey of RAIN suggests building the "perfect persona" by storyboarding the values of your brand and how those tenets could be represented vocally. "Can the audience relate to (the voice) or does the VUI sound like it's the audience's child or parent (and if so, is that a good or bad thing?) Selecting a persona is just as important as selecting a brand ambassador and should go through just as stringent a vetting process."



**Elissa Dailey**  
Senior Strategist  
RAIN  
@elidai

**"Your voice user interface needs to be the literal voice of your brand. It should speak with confidence and authority, keep responses brief, and provide the right amount of information."**

## 4

### Keep your brand's voice authentic and human

A challenge just as prominent as that of selecting a persona, is appropriately communicating and maintaining brand voice and tone. Daily adds, "Similar to how you inject your brand voice into messaging, marketing, and other communications, you will need to appropriately script your desired voice, and corresponding tonal adjustments, into your dialogue flow - and on top of that, will need to make it feel authentic and natural."

An added layer of complication is that while many marketing channels are static in their messaging, voice is not, and requires you to account for what can be hundreds or thousands of interactions all maintaining a consistent voice.

Since we already know that people aren't always comfortable with conversing with machines, it's essential that the voice sounds natural and ordinary. The conversation should flow naturally. Forcing the user to press the button every time they talk can be cumbersome and unnatural.

41%

of people who own a voice-activated speaker say it feels like talking to a friend of another person.

Source: [Google](#)

Golembiewski, an expert in both conversational design and product design, often speaks on this topic with other creators of voice user interfaces. "The VUI's discovery, usability, and user satisfaction increase when it has a consistent voice that modulates based on the situation and users' emotional states along with a well-defined persona," she advises. "Designers can build upon the established voice, tone, and persona by refining word-usage rules and writing style guides. It is important to establish writing examples for the VUI when documenting brand guidelines."

Dailey agrees. "If your scripted dialogue isn't something you'd say out loud, consider adjusting," she said. "You can test this by "reading lines" with a partner. Dailey also recommends further anthropomorphizing your brand voice with words and phrases like "thank you," "got it," and "please" to personalize conversations and speak to the user in a way they're used to from talking with other humans.

**"The VUI's discovery, usability, and user satisfaction increase when it has a consistent voice that modulates based on the situation and users' emotional states along with a well-defined persona."**



Lauren Golembiewski  
CEO  
Voxable  
@LaurenGolem

"Scripts are a key component of building a voice experience, as they quite literally guide the experience from point A to point B. Just like in any other kind of script (movies, TV, plays, etc), character consistency and development is important to ensuring fluidity and can increase the potential for a deeper connection with your audience." adds Dailey. "To maintain a consistent voice throughout, try building a character profile to define limitations of voice and then account for areas of tonal fluctuations throughout the script. Outline, via a storyboard, the values of your brand and how you think those could be communicated or represented, the sense or mood you want to put out as well as get back from your users, and more. Additionally, storyboard out a variety of interactions you expect to have with your users to determine the specific tonal adjustments you'll want to make against each nuanced situation."

## **5 Anthropomorphism**

Dailey advises that in order to imbue some of the same characteristics that we as human beings maintain when conversing, it is critically important to

adapt a variety of anthropomorphic qualities in your voice experience, such as:



**Natural, authentic language:**

Write scripts the way you'd speak naturally. If your scripted dialogue isn't something you'd say out loud, but rather something you'd read, consider adjusting. You can test this by "reading lines" with a partner.



**Contextual relevance:**

Pick conversations back up right where they left off to show the user that you've not forgotten the progress they've made. This creates a level of understanding for the user that each conversation is built off past conversations, growing the relationship between user and voice experience.



**Conversational markers:**

Use words and phrases like "Thank you," "got it," and "please" to personalize the conversation and speak to the user in a way they're used to from conversations with other humans.

**"In order to imbue some of the same characteristics that we as human beings do/maintain when conversing, it is critically important to adapt a variety of anthropomorphic qualities in your voice experience."**



**Elissa Dailey**  
Senior Strategist  
RAIN  
@elidai

**Here are some examples of how RAIN helped their clients solve for this:**

- RAIN worked with Sesame Street to recreate Elmo's voice in a kid-targeted voice experience, ensuring they used the appropriate

language, voice, and tonal adjustments to ensure an authentic Elmo and Sesame Street feel. RAIN took thousands of clips and strung them together to form scripted Elmo statements and responses, accounting for a variety of scenarios in which Elmo may have to react.

- For Starbucks, in order to incorporate within the Alexa Skill some of the in-person characteristics of a barista that make the ordering experience successful and unique, the RAIN team visited brick and mortar locations to listen to how the baristas spoke. This helped them craft the voice and language of the experience, understand the importance of using a customer's name, and more.

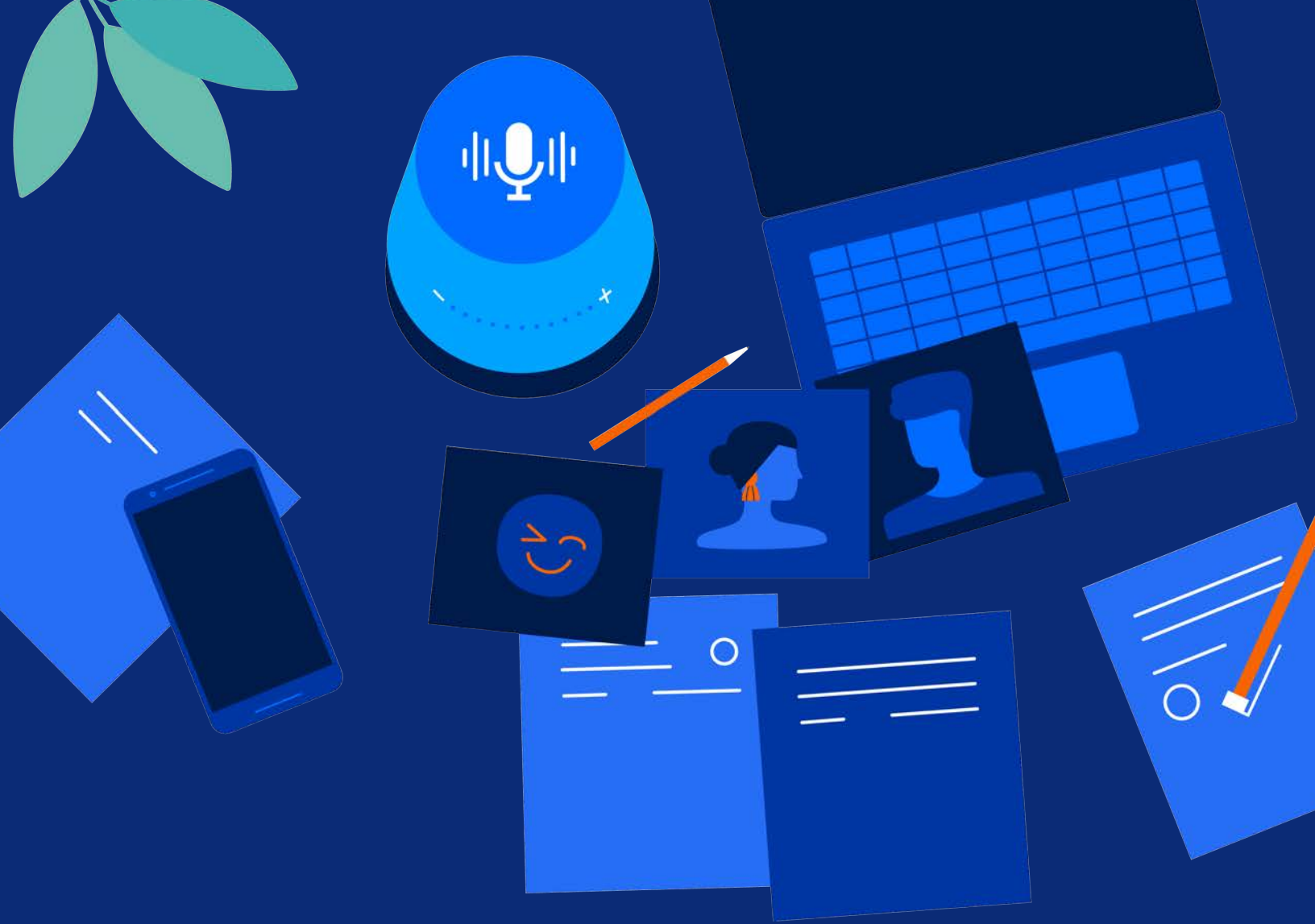
## **User Research and Testing Advice**

Dailey of RAIN shares a couple of ways to test the persona, voice, tonal adjustments, and anthropomorphic characteristics you've selected for your voice-based experience. "The simplest way to fix something that can occur early on in your voice design process is a targeted survey to, or focus group with, identified audiences." She adds that a survey can help gauge immediate reactions as well as can uncover unexpected insights for your consideration prior to actually designing and developing a prototype.

"Another way to test these elements is through user-testing of a low-fidelity prototype." Having users log their interactions and reactions to those interactions can help you get a sense of how well the experience is providing value and connecting on an emotional level, if at all.

In our next chapter, learn more about infusing personality and humor into your VUI. With the right amount of wit, banter, and ease, your brand voice can be much more human.





Chapter 02:

# Infusing Personality and Humor

# How Voice Assistants Win Friends and Influence People

Since the earliest science-fiction movies and books, talking computers and robots were portrayed as something to fear. From the ominous-sounding “I can’t do that, Dave” from 2001: *A Space Odyssey* to the trope of the countdown to self-destruction (always stopping at 1), computers have been the perennial bad guy. Now that our devices really do talk, it’s vital that we help them shed their reputation. The best way to do that is to create a nicer, friendlier, trustworthy voice user interface (VUI). More than anything, what people want from their VUI is a more human experience designed to help them as a friend would.

We may not notice it, but as people, we look for characteristics and personality in nearly everything we interact with. In fact, we rely on speech to evaluate people on honesty, trustworthiness, overall intelligence and education (knowledge). That’s why it’s important for brands to always keep human nature in mind when designing a VUI. Even if you don’t focus on building a specific personality for your VUI, having one is inevitable—your users will naturally personify it.

People dislike talking to cold and calculating robots and as a result, many innovative brands are already being more thoughtful when it comes to their voice AI’s personality.

*Even if you don’t focus on building a specific personality for your VUI, having one is inevitable—your users will naturally personify it.*

Defining your personality should be high on your list when you start planning to build conversational experiences for your customers. It needs well-written dialogue, small talk, empathy, humor, and emotional intelligence that caters specifically to your users. The actual voice and tone of the conversation should be in sync with your brand voice, too—and that is one of the biggest hurdles in designing a VUI successfully.

Interacting with a voice assistant should be a seamless, frictionless conversation that’s comfortable and simple. These best practices can help guide you as you build your voice interface.



# Best practices:

## 1 Craft the right personality for your brand, and for your audience

According to Will Hall, Chief Creative Officer at RAIN, “One of the biggest challenges with voice assistants is that they sound like humans... almost.” The almost problem is more profound than it may at first seem, though. Because the voices are almost human, we as users tend to graft human expectations onto these interfaces that simply aren't yet up to the task of conversing intuitively, understanding emotions or building true empathy. As architects of these experiences, we must find ways to set appropriate user expectations and bridge the gap with thoughtful design.



How do you want your VUI to represent your brand? If you have a company mission statement, that's a good place to start. What values do you want to convey? Once you have that guidance, use your imagination. If you were to personify your VUI as the ideal employee for your company, how should that employee sound? How do they carry themselves? What's that person's attitude when talking to customers? What kinds of words do they use?



**Will Hall**  
Chief Creative Officer  
RAIN  
@RainAgency


**"Because the voices are almost human, we as users tend to graft human expectations onto these interfaces that simply aren't yet up to the task of conversing intuitively, understanding emotions or building true empathy."**

Next, turn your attention to the users. Defining your customers' identity is the first step toward building your VUI's personality. Write down what you think this voice assistant's job description is. Answering simple questions like, "Are our clients professional or leisure users?" and defining their age range, gender, geographic spread, and language can help you build that job description to meet the needs of customers better. When you build this job description and understand its purpose and audience, creating its personality will be much easier. For example, if the voice assistant provides account information, it could be more formal and knowledgeable, but if it offers kids' entertainment, it should be friendly, funny, and use age-range appropriate dialogue.

### **Thinking about your VUI's gender**

Choosing your VUI's gender can be tricky. Like everything in life, you should put this to the test and see how people react. Start by simulating an interaction and note what the users' expectations are with regard to gender. "While you don't want to foster gender stereotypes, think carefully about whether your company's voice is best represented with a male, female, or a [genderless voice](#)." Opinions vary about what's best of course.

One school of thought urges voice designers to challenge gender stereotypes. IBM's Jeopardy-playing AI Watson speaks with authority and a male voice. But Alexa, built to be helpful and supportive, is female. Apple's Siri and Google Assistant offer users the opportunity to choose the preferred voice, gender, and accent of their AI, but that choice sometimes runs the risk of diluting brand recognition.




## Acapela TTS Voice Collection


By Acapela Group

---


**Description**  
 This domain provides voice output for other domains.  
 To use one of these voices, set the "ResponseAudioVoice" field of the request info to one of the voice names below and set the "ResponseAudioShortOrLong" field of the request info to either "Short" or "Long". Then, whenever a match occurs in any other domain, the result JSON will include the voice data in the "ResponseAudioBytes" field of the CommandResult JSON and the "ResponseAudioBytes" field of each dynamic response JSON.  
 The voice data will be Base64 encoded to form a string. The raw data bytes before Base64 encoding will be in WAV format.



**Display Screen Not Required**  
 This domain does not require the client to have a display screen to be fully functional.





**Client Integration Not Required**  
 This domain does not require the client to write additional code to work correctly.



**Credits**  
 This domain costs 0 - 0.25 credits per query. For more information on how credits work, check the [Pricing](#) page.

**Output Voices**  
 The table below lists the available voices provided by this domain.

Voice Name	Sex	Languages	Sample
Sharon	female	American English	
	Description : The phonemes for this voice can be found <a href="#">here</a> . More detailed documentation can be found <a href="#">here</a> .		
Lily	female	American English	
	Description : The phonemes for this voice can be found <a href="#">here</a> . More detailed documentation can be found <a href="#">here</a> .		

You can [explore and play with voices](#) provided by Acapela on the Houndify site. Try out some different voices and think about your brand. Which seems like it might be a good fit for you?

Lauren Golembiewski, CEO and Co-founder of Voxable advises clients to be ready for controversy no matter which gender you choose to represent your brand. "The intersection of a social issue and technology is always going to be rough," she said. "The best way to be prepared for it is to have a good answer for why you chose the gender you did. Whatever your decision, if it ties into your core mission authentically, people won't judge."

Most importantly, design your VUI's personality at the very beginning of the conceptual stages with well-written copy that caters specifically to the target user. That's why it's so important to create a persona that resonates with your audience and also with your brand.

**"The intersection of a social issue and technology is always going to be rough. The best way to be prepared for it is to have a good answer for why you chose the gender you did."**



Lauren Golembiewski  
CEO  
Voxable  
Twitter @LaurenGolem

At Soundhound Inc., Director of Product Design Michael Mahmood advises taking a holistic approach when creating voice experiences. "Know your audience, understand their motivations, and develop a deep sense of empathy for their pain points when it comes to using voice. Once the user persona is clearly defined work tightly with marketing to ensure the tone and communication style aligns with the overall branding strategy. Also, work closely with the engineering teams to make sure different uses cases are accounted for as voice and speech can get very complicated."

The SoundHound design team is continuously learning and listening to their users as the voice revolution moves ahead. "We at Soundhound Inc. know we're just scratching the surface when it comes to the potential power of voice, and we're excited to lead the way with meaningful voice-powered experiences for our customers."

## **2 VUIs break the ice with small talk, then win customers over with empathy**

"If there's anything that will connect your users with your voice assistant, it's small talk," advises Rachel Batish, vice president of product at [Audioburst](#) and author of [Voicebot and Chatbot Design](#). "Either as side-interactions or as part of the main discourse, it's the small talk that people will talk about and remember." Human-like chatter is a feature that both demystifies AI and warms people to it. Users become aware that they can simply talk normally to a VUI, and eventually come to depend on it. But VUI, like any technology, isn't perfect, so designing humorous or witty quips and come-backs are valuable in keeping the user smiling when the system can't deliver the desired information or action. "The more chatter sentences and events you

put into your VUI the better it is," Batish said. "While it's not a huge investment for you, it makes a huge difference for the user."

SoundHound Inc. worked closely with writers on the marketing team and found that people like a little bit more fun than robotic responses. We found that being friendly, casual but polite, and not too verbose goes a long way. Mahmood is inspired by tiny touches used by other companies that keep the user encouraged even if the system can't provide instant gratification. "Think of that little dinosaur you sometimes see in Google. It's telling you it can't find what you're looking for, but you don't mind because it's so cute. You just try again. That's what we aim to do with voice. We want to make even a bad experience delightful. That helps users to be a little more forgiving. We don't want to discourage them from getting better at (using voice assistants). Of course, we also want to be helpful. We try to anticipate users' needs. We might provide an answer and a follow-up question like: "Is this what you're looking for?"

**"A VUI, like any technology, isn't perfect. Designing humorous or witty quips and come-backs are valuable in keeping the user smiling when the system can't deliver the desired information or action."**



**Rachel Batish**  
VP of Product  
Audioburst  
🐦 @rachelbatish

### **Empathy without losing brevity**

While it is important for a VUI to be approachable from the very beginning, it should not be repetitive. It's important to provide your VUI with the ability to anticipate moments and keep a conversation going during the interaction flow with users. It's essential to add variety to its language and shape its ability to retain information so it can constantly adapt to the user's requests.

In person-to-person interactions, building up a sense of empathy is essential

in establishing a proper and well-meaning relationship. For the sake of your user base, your VUI should be no different. In many cases, a general tone of positivity and affirmation goes a long way in keeping your users engaged even in worst-case scenarios. For instance, if the VUI is unable to fulfill a user's request, it should take full ownership for the lack of understanding and then follow up with corrective suggestions. This helps to build confidence and trust with the user, encouraging additional interactions and queries because there is now an increased belief that the outcome will be beneficial versus having no value.

Also, remember to have your VUI remain sensitive to the user's answers. Adjust subsequent options based on how your user answered previously, anticipate particular errors based on functionalities that your VUI does not have, and acknowledge any difficulties within the conversation flow to encourage the user's patience and perseverance.

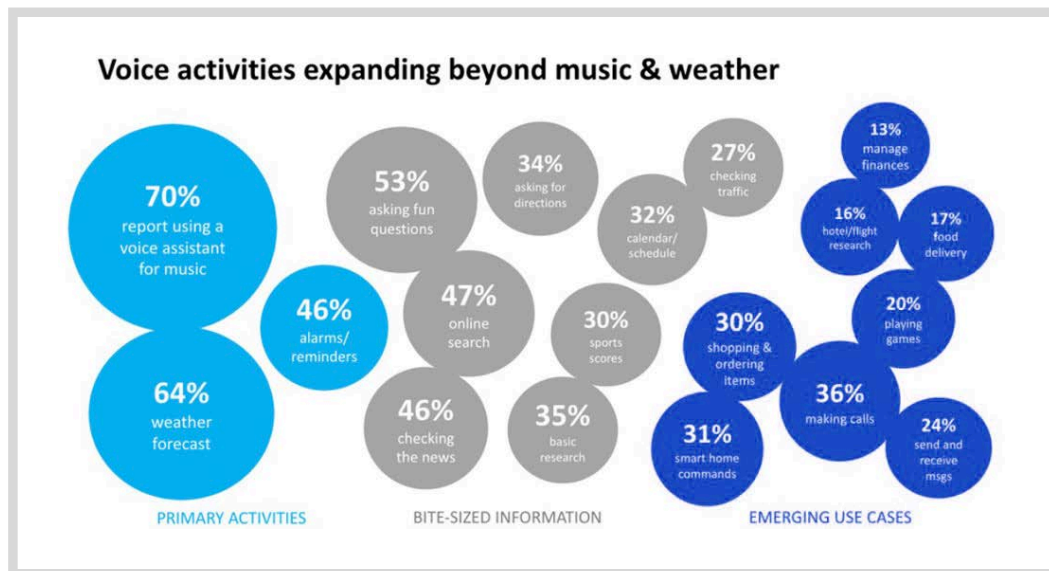
### **3 Never underestimate the power of humor**

At the very base of a VUI is the way it speaks and responds to the user. The more personable its tone, the more it helps to increase brand affinity. When it comes to writing the dialogue, it's best to opt for the most "human" interactions possible. This means including occasional small talk as well as humor and even general reference to pop culture among the VUI's many other functionalities. Jokes or witty remarks help the user to become more comfortable and acquainted with speaking and interacting with the VUI, potentially increasing further use and engagement. In fact, according to a study by Adobe Digital Insights, 53% of U.S. smart speaker owners use their smart speaker to ask "fun questions."

**53%**

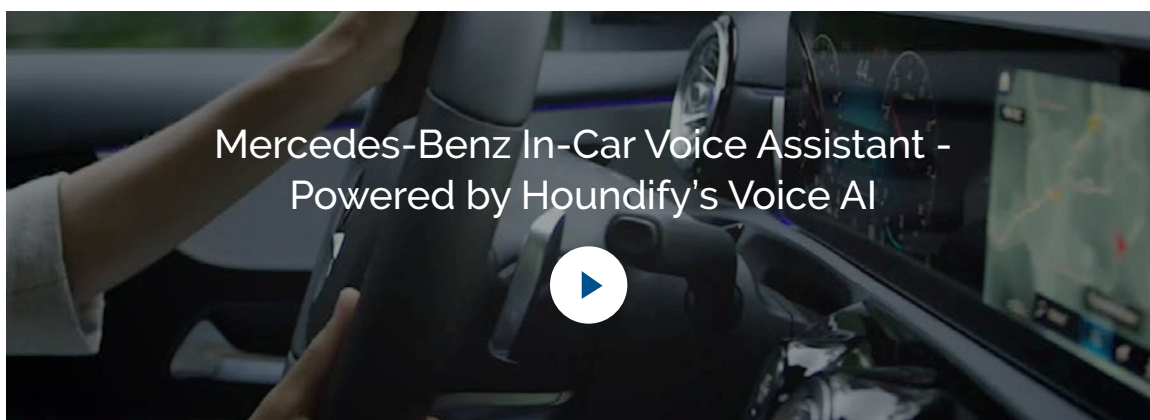
of people who own a voice-activated speaker say it feels like talking to a friend or another person.

The Mercedes-Benz infotainment system MBUX's ability to crack a joke every now and then allows it to be more entertaining and "human." Senior Engineer Mihai Antonescu advises brands to "never assume that you have enough utterances as that's a mistake that you can easily make in the beginning. Don't just think about "How would I ask this?" Because if you ask your friend, he will probably ask it very differently. So keep asking people about how they would ask for one specific intent and then constantly try to improve your language model," he adds.



Source: [Adobe Digital Insights](#)

You can also take this small clip of Mercedes-Benz's MBUX voice system for instance. Providing occasional moments such as these throughout the VUI experience encourages users to be more forgiving during moments in which the interface is unable to complete tasks or answer certain questions. The power of natural, friendly conversation should not be underestimated.





Voxable's Golembiewski says that it's always better to be forthcoming with your audience. For example if a user is talking to a chatbot, let them know. If users are aware they're not interacting with a human, they're likely to be much more receptive to the mistakes a VUI might make if they weren't expecting a human interaction. "Humans personify and build empathy with inanimate objects all the time," Golembiewski adds. "Think about all the people who name their cars!"

**"Never assume that you have enough utterances. Don't just think about 'How would I ask this?' Because if you ask your friend, he will probably ask it very differently."**



**Mihai Antonescu**  
Senior Engineer  
Mercedes-Benz R&D  
North America  
Twitter @MBRDNA

## 4 Additional expert tips

Here are the top 3 things that RAIN's chief creative officer Will Hall recommends:



### **Mind the Gaps:**

In voice experiences, users insert their emotions most into the negative spaces of conversations. For example, if a consumer used a customer service skill that had long pauses after each question, the user will channel "are they even listening?!" into that gap. Understanding that these moments will spark the most emotion means that we must anticipate and respond accordingly.



### **Anticipate:**

We don't always say exactly what we mean. As designers, we must anticipate and predict intent wherever possible. For example, if a user checks their balance with a banking app, they might just want to know if they can purchase something instead of



the actual available funds. If a user asks Alexa about the weather, they might want to know whether they should bring an umbrella or not, as opposed to the actual temperature. The more we can enable the tech to intuit intent and anticipate requests, the more frictionless and empathetic our experiences will become.



**Listen:**

We think of voice as a “focus group of one” because of the specific nature of conversation. Users will tell you specifically what they want in voice. For example, with Tide we created a Skill that accounted for a hundred or so stain “recipes.” By listening to requests, though, we found that nearly 20% of the things that users asked Tide for were things that they don't yet have products for. This is a powerful truth that can help voice experiences drive personalization for users and glean qualitatively rich insights for brands.



**Will Hall**  
Chief Creative Officer  
RAIN  
🐦 @RainAgency

**"As designers, we must anticipate and predict intent wherever possible. The more we can enable the tech to intuit intent and anticipate requests, the more frictionless and empathetic our experiences will become."**

In our next chapter, learn more about enabling more natural human-VUI interactions. Your VUI needs to understand context, location and just the right amount of user info to be a great conversationalist.



Chapter 03:

# Enabling Natural Interactions



# A Little More Conversation

## Crafting Genuinely Meaningful Experiences

Intuitive interactions depend on a solid design process stemming from user research and usability testing. Users can say anything to the VUI, so designers must gain a deep understanding of users' needs to influence their design. A VUI should be able to anticipate what users will say and the contexts that are most important to them. Once the user interface is implemented, usability testing is vital in determining where the VUI may fail at handling users' requests.

The irony of enabling natural interactions is that the biggest barrier to achieving it is technology itself. Though the temptation to speak to a bot in a series of keywords is great, it's not how people talk, and it defeats the point of creating a natural

language between people and machines. The minute a VUI breaks away from human-style interaction, it's asking people to change behaviors, and that creates the very friction a VUI is designed to erase.

Everyone speaks just a little differently, so VUI designers have to think of all the different ways a person might ask a question. As Heidi Culbertson, founder and CEO of [Marvee](#), a company dedicated to voice design for older adults said, "Expect the unexpected, and then expect to be surprised that you didn't expect something." Voice-enabled software starts with the end user, so continual research, user testing, and iteration to improve the language model is key.



**Heidi Culbertson**  
Founder & CEO  
Marvee  
🐦 @hcultbertson

**"Expect the unexpected, and then expect to be surprised that you didn't expect something. Voice-enabled software starts with the end user, so continuous user testing, and iteration to improve the language model is key."**

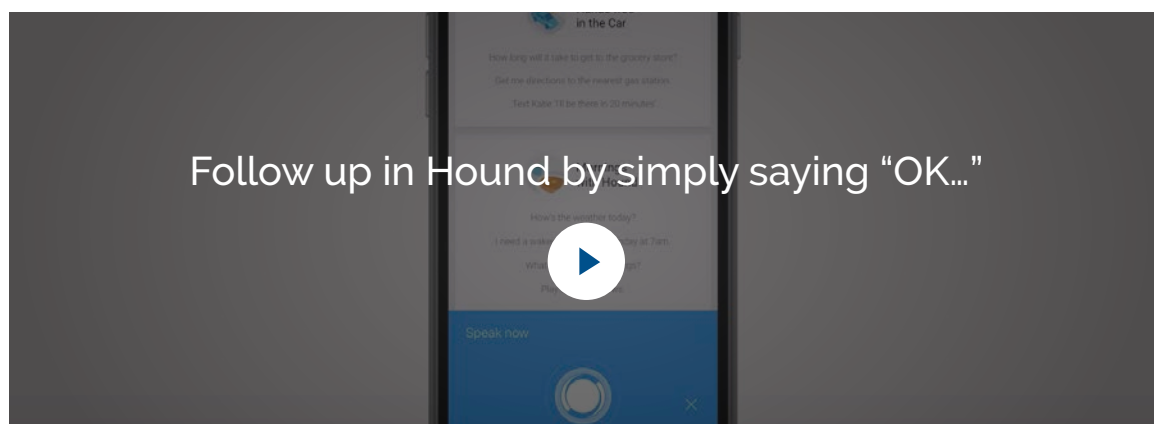
# Best practices:

## 1 Create voice experiences that are context-aware

Think about late-night television show hosts. They're masters of conversation. They have a few minutes with each guest to ask relevant questions, get the answers the audience wants to hear, and to ask follow-up questions that make sense in context with the guest's previous responses. Since their time is limited, they have to ensure the questions don't require more brainwork than the guest or the audience can process in that time. A well-designed VUI should work the same way.

*How the VUI responds is the make-or-break moment for a user. This is when they will learn if they can trust your product to hold up your end of the deal.*

The VUI should have the ability to provide facts that the user might find useful and still stay relevant and sensitive to context. For instance, if the user asks for the weather or news nearly every day, have your VUI suggest those domains more frequently when prompted. Through this design, you can establish a proper relationship between your interface and the user. The VUI understands what the user wants and is able to adapt to those wants, creating a small but shared hub of memory and conversations.



Mihai Antonescu, a senior engineer at Mercedes-Benz R&D, has been working on [MBUX](#), an in-car voice experience powered by Houndify. It's his team's job to make sure the car and driver can get along without distraction and provide the luxury experience inherent to the Mercedes brand. "Our biggest learning is that context matters," Antonescu said in a [2019 Houndify webinar](#). "In our case, the customer in the car is probably driving. He has a lot of questions about the car, so we make sure that the VUI can answer all the questions in that context. Then, we expand on new things, and see where the context takes us. When the users says "Hey Mercedes," the MBUX responds to natural-speech voice queries and commands that range from the weather to nearby restaurants, to features like turning on the air conditioning, remembering the driver's preferred temperature settings and rolling up the windows."



How the VUI responds is the make-or-break moment for a user. This is when they will learn if they can trust your product to hold up your end of the deal. Did the voice assistant understand your request, and acknowledge understanding? Can it handle multi-part questions like "Find me Asian restaurants within 10 miles that are not Chinese, have at least 3 stars and are open late," that require access to multiple domains? Did it provide a short, relevant response and the opportunity for follow-up questions?

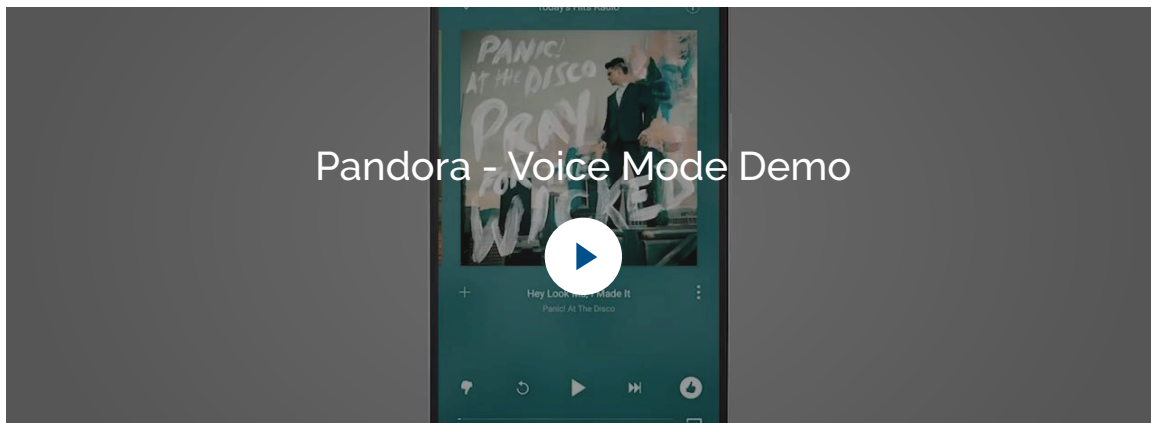
At Pandora, which is known for its proprietary algorithms that help the brand personalize its music-streaming service, context was just as important for its voice assistant. "Music is one of those contexts where I don't have to think,"

said Ananya Sharan, product manager for Pandora's voice mode. "Sometimes I just want Pandora to know what I like. I want to say 'Play me something awesome,'" and it does. Thinking of the context really helped us to prioritize what we wanted to build and how we wanted to design the interaction."

**"Sometimes I just want Pandora to know what I like. I want to say, "Play me something awesome," and it does. Thinking of the context really helped us to prioritize what we wanted to build and how we wanted to design the interaction."**



**Ananya Sharan**  
Product Manager  
Pandora - Voice Mode  
🐦 @\_ananya\_



Ben Steele of RAIN feels the threshold for natural conversations with our voice assistants is a long way off, even with recent notable tech innovations. "We're able to design a pretty frictionless experience, but what we're really looking forward to is for our assistants to interact on our level instead of us on theirs, right? Still, there are plenty of best practices that can reduce friction and improve the perception that you and your voice assistant are in an ongoing contextualized conversation"



**Ben Steele**  
Producer  
RAIN

 @RainAgency

**“We’re able to design a pretty frictionless experience, but what we’re really looking forward to is for our assistants to interact on our level instead of us on theirs, right?”**

The best way to ensure context and personalization is through data collection. And the best way to collect data is through usage. The more volume of data you get, the more signals from customers you get that can help optimize the experience. Whether it's gradually learning your musical tastes and what you listen to at various times of the day or remembering your driving habits and temperature preferences, machine learning and AI can automate your favorites and deliver them right to you.

A cool feature in the MBUX voice assistant is that you can just say “Hey Mercedes” and mention the state that you’re in and the system will recommend things for you. Antonesco explains: “I’ll give you a couple of examples: Let’s say that you’re hungry, or feeling tired or cold. You can just say “Hey Mercedes, I’m hungry” and the system will recommend restaurants because you’ll probably want to look for food in the near future, and this is related to the state you just communicated to the assistant.”

## **2 Consider how different people speak**

Heidi Culbertson founded a voice design company to help her aging parents navigate the world of technology to improve their independence, and in doing so, became an expert on creating VUIs that are more accessible and cater to different audiences. Her specialty is creating voice experiences for older audiences, but paying attention to the needs of the users is important to any age group of users. “As you’re designing a VUI, you need to really know who your audience is and start to personalize experiences. It’s matching how they’re comfortable speaking,” Culbertson explains. “The length of phrases, their word choices, their cadence, their rate of speech. VUI design is truly about the listener.”

Antonescu agrees. "Older people (tend to) ask full questions, and then they say "please search for this." That's very different from the way younger generations search for something. That was breaking our model, so we had to adjust it so we can handle things like 'thank you' and 'please.' One mistake voice designers make at the beginning of the process is to assume that they've written enough utterances. "You think, 'how would I ask this?'" Antonescu said, "And then you ask a friend and he'll ask the same question very differently. So keep asking people about how they would ask for one specific intent, and then constantly try to improve your language model."

5%

**of voice assistant users find them effective at executing commands every time**



63%

**of them find it takes two or more attempts to get the results they want**



Source: [Zazzle Media - 2018 State of Voice Search Report](#)

The need for personalization based on audience reaches across verticals. Sharan remembers working on Pandora's language model with expectations that most people would say 'Hey, Pandora, play me the xxx song.' When it came to user testing, she was surprised. "People were asking for it by lyrics," she said. "People were singing lyrics to the voice assistant. There were kids talking and other noise in the background. So we had to adapt and react to all of that. One solution was to clarify with the user what they meant and present them with options. But how much conversation do you engage with so that it doesn't start to feel like we're insulting the user?"



"Over time, remove as many barriers to get to the point as reasonable, but be aware that you're closing doors when you do so. Weigh the value of an efficient experience against the constraints it places on the user, and offer avenues for them to remove or reconsider those constraints" Steele adds. "In general, be easier than a mobile app, then once you are, wean your users off their screens by facilitating habits around voice."

## **Top VUI best practices from Mercedes, Pandora, and Marquee**

### **DO:**

- Keep your focus on your top use cases, and don't try to be everything for everyone
- Read your voice assistant dialogue scripts out loud
- Start testing at the beginning of your project, not after you've already scripted everything
- Remember the importance of context
- Understand that personalization is absolutely number one—it's how you differentiate yourself

### **DON'T:**

- Believe your assumptions are correct
- Assume that the user is hearing what you think you're saying
- Allow script-testers to face each other—that allows them to take in body cues and hand motions that won't be available in a voice-only environment
- Add more information to an answer than a user can absorb

## **3**

### **Mom was right. Politeness counts**

It goes without saying that voice assistants need to be polite, but people, cultures, and languages have different ideas about the concept of

"politeness." Linguist Julie Belião, Director of Quality at [Unbabel](#), thinks about this question daily for the company, which provides AI-powered multi-lingual customer support and service. She follows nine specific rules that define politeness and civility around the world. "You need to adjust the politeness level and reply to your customers in a way that shows you care about them," she explains.

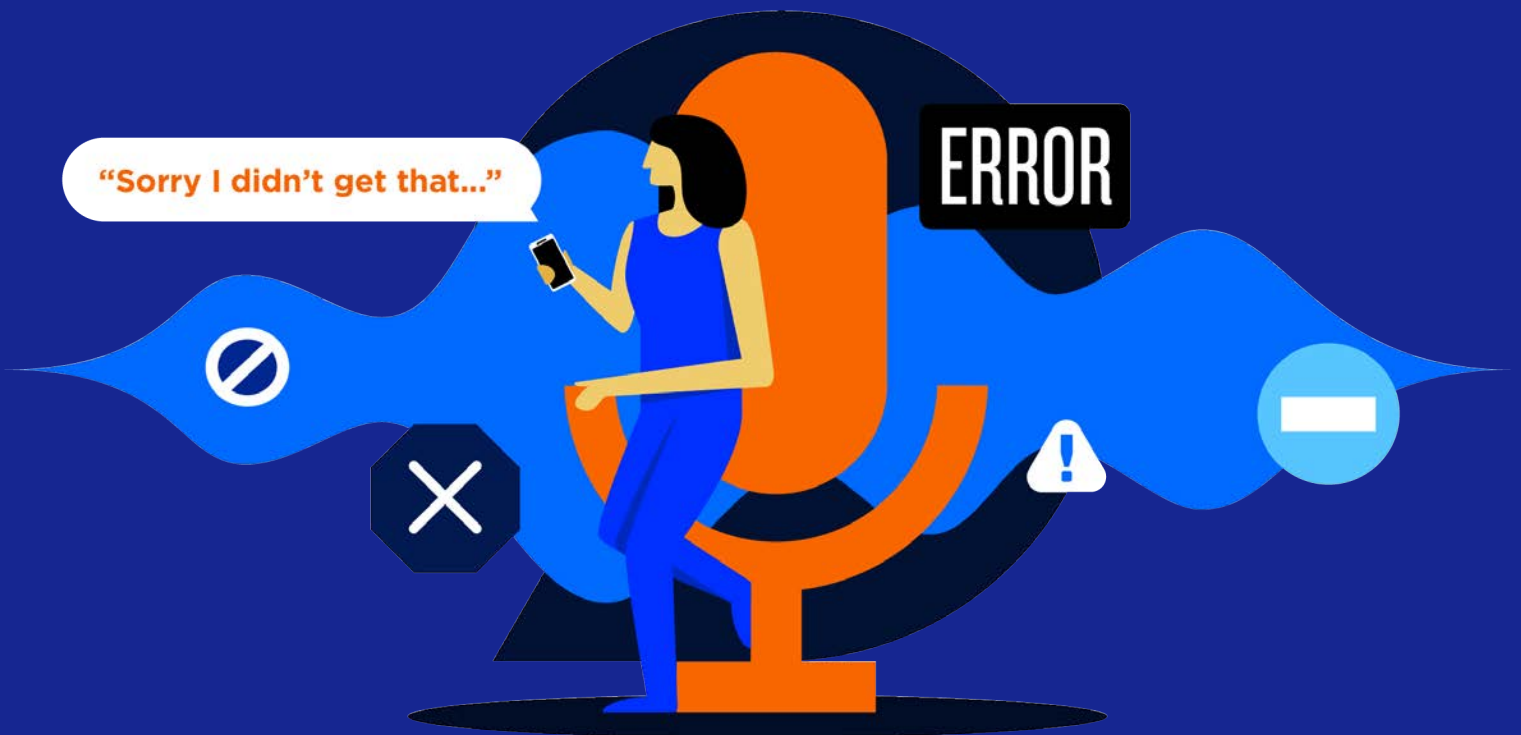
Her guidelines include tips like using softer language, rather than being direct. "Being very factual can sometimes sound too direct or give the wrong expectations," she said. "So you can add vague expressions like *about*, *kind of*, *sort of*, *stuff*, and *things*." Another way to soften the tone (and seem more polite) is to use modal expressions (*could*, *might*, *should* and *would*, for example). Use positive, affirming words to come across as assuring and energetic." Good choices to add include words like *amazing*, *awesome*, *great*, *good*, *gladly*, *definitely*, *absolutely*, and *understand* to make a customer more comfortable and reassured.

In the next chapter, learn the best practices for user prompting and error recovery. It's hard to follow a path you can't 'see,' or in this case 'hear.' Learn how prompts and error processing can help improve user experience.

---

## Chapter 04:

# Prompting and Error Recovery



# Keeping the Conversation Going

For the most part, voice user interface (VUI) is an audio-only experience. That being said, designers need to key into the flow of two-way conversations and create ways for the assistant to help the human user navigate, find things, and most of all, feel heard. Assistants can often give more information than the human can absorb, but at the same time, users need enough information to complete their task.

Though some smart speakers like Amazon Echo and Google Assistant only have voice interfaces, the visual component to the app versions of each and any VUI is of utmost importance too. It's easy to forget the visual interface when you're working on voice technology. But visual cues help users understand what's happening, and what they need to do to take action.

Since AI has limited parameters and humans have nearly infinite ways of conversing, count on some misunderstandings between the two. This doesn't have to be the end of a budding relationship. If your VUI misunderstands or sends users down an irrelevant path, look for ways to both recover, and for the voice assistant to learn from its mistakes.

Then, there's the issue of the users themselves. RAIN Producer Ben Steele

explained it best: "I don't know what to say," he confessed. "I'm bad at talking to them. I stutter over any command I haven't repeatedly used before to the point where even another human being wouldn't be able to derive a coherent sentence out of it. I'm just not able to process my side of the interaction in the same way that I would with another person." He adds, "I have a limited window into things it can do based on what I've needed to get it to do in the past and maybe a few things I've derived from commercials. Even beyond the features, I don't really know how I should structure my sentences and I don't know if I mess up my sentence if I need to just give up and start again or push through."

In the same vein, scripting for voice assistants must be carefully crafted so the human users learn how to interact with it. Steele adds "I don't really know what phrases are attached to functionality even if my words are understood, so I'm just sitting there trying to shout keywords at it. Me and my voice assistant, we're not on the same wavelength."

In this chapter, we'll discuss how to help machines and humans communicate better.

# Best practices:

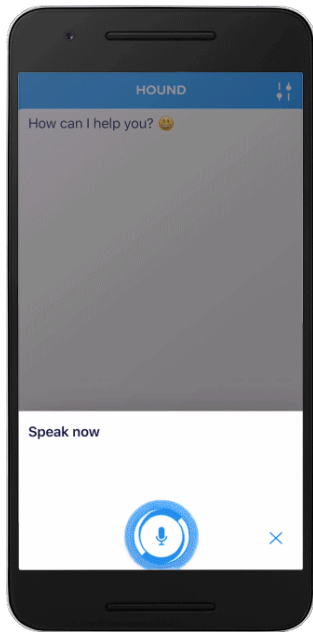
## 1 Give users confirmation and a subtle roadmap

How do you know when the VUI is listening? How do users know how to summon their voice assistant if they don't want to call its name? What cues do people need to know when to talk and when to listen? The listening screen conveys that the app is focused on getting user input through voice — not unlike when the keyboard slides up to get user input through text. This screen usually appears in the form of an overlay and can take up as much real estate as you feel is necessary. The listening screen should contain several key elements that help bring the voice and visual experience together.



The first consideration is a microphone button or other part of the speaker that lights up as an easy-to-spot indication that the user is being heard. Siri's multi-colored sound waves show users that Siri is listening. The animation has the same purpose as Alexa's blue ring of light or Hound's twirling blue ring.

It's also helpful to visually represent the audio input because it provides a stronger sense of connection between the user's voice and the product's understanding. Common visualizations include animated sound waves and level bars. And show users the way out. Provide a way to turn off the listening screen at any time.



The audio cues are just as important as the visual ones so users can tell that the device is listening even if they're not looking at the device.

Here's some advice from Octavio Menocal, senior voice experience engineer at RAIN: "Include in the reprompt the type of phrases you expect from the user in the context. If users respond with an unexpected intent, catch the failure, and send a response including the type of phrases your app expects to move forward."

### Here's an example of a taxi voice skill in Alexa:

---

**Alexa:** Welcome to taxi driver. I can contact a taxi to drive you to your destination, or give you a quote, which would you like?

**User:** I want to go to the airport (Unexpected intent)

**Alexa:** Excellent! I just need to know if you want me to contact a driver, or if you simply need a quote, what's your interest?

**User:** Oh right, I want you to find a taxi for me.

**Alexa:** Great! You told me you wanted to go to the airport, correct?

**User:** Correct -> (User is pleased the app remembered his destination and didn't ask for it again)

**Alexa:** Ok, now, where would you like the driver to pick you up?

(User does not remember and asks her mom next to her, while talking to Alexa)

**User:** Mom, what's the name of this street?

**Mom:** William Street, 84404 (Yes! This dialog happened at the time user was talking to Alexa, and guess what, Alexa heard: *"What's the name of the street we will be at, at 4pm"* (Unexpected intent))

**Alexa:** Sorry, I didn't understand, where would you like the driver to pick you up? (We didn't stop the conversation, instead we handled the FallbackIntent and returned a valid reprompt)

**User:** William Street, 84404

**Alexa:** Ok, give me a second, I'm looking for the closest drivers -> This is a progressive response while your app gets information from your external API

**Alexa:** Ok, I found 5 drivers available, the ETA to pick you up is 5 minutes, the price is \$40, do you want me to contact the closest driver right now?

**User:** Yes

**Alexa:** Excellent, the driver confirmed he's coming in 5 minutes. I just sent you an email with the receipt of your order. Thanks for using taxi driver, have a nice trip!

## 2 Create a strong strategy for error recovery

"Human conversation is naturally replete with errors," says Lauren Golembiewski, CEO and co-founder at [Voxable](#). "Through experience, most human brains learn how to correct conversational errors fairly seamlessly. On the other hand, VUIs need to be encoded with extensive error recovery." Golembiewski cites an example: "What happens when a user says something unexpected to a VUI prompt? How should the system respond if a user says 'I don't know' to the VUI prompt: 'Are you sure you would like to transfer \$2000 to that account?'" This error recovery is as simple as considering what might happen in these kinds of scenarios. But it can get more complex depending on the risk involved in the interaction.

At [Marvee](#), CEO and VUI designer Heidi Culbertson depends on analytics to help her get to know her audience and understand how they think. "The better you know the actual user, the better your error management will be," she says. "Error management is literally taking it almost word for word and designing your VUI so that you don't end up in an endless loop." Another pitfall with errors is the possibility of losing the audience. "You lose retention when it's not an easy experience," she explains.

**"Don't stop the conversation when an error occurs. If your app has a syntax error or maybe an operation with an external resource fails, return a response informing the user something unexpected happened and they can come back and try again after you have fixed the error."**

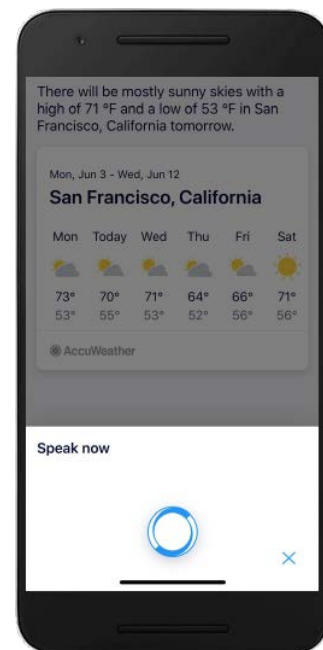


**Octavio Menocal**  
Sr. Voice Experience Engineer  
RAIN  
🐦 @omenocalc

"Don't stop the conversation when an error occurs," Menocal adds. "Some frameworks offer a nice handler to catch an unexpected error in the code. If your app has a syntax error or maybe an operation with an external resource fails, return a response informing the user something unexpected happened and they can come back and try again after you have fixed the error. You can also print that error in your server, or send it to your email, Slack or via SMS so you get alerts of what's wrong and think of how you can fix it quickly."

### 3 Help keep users from getting tongue-tied

SoundHound Inc. has been designing voice-enabled AI technologies for 14+ years. If there's one thing we've learned from implementing voice experiences, it's that users usually aren't sure what to say. Sometimes users don't know that they should say something when the listening screen appears. Even though they have initiated this screen with the tap of a microphone button, they are often caught off-guard and can feel like they are being put on the spot. In fact, our data shows that many first-time experiences are met with silence.



Prompt asking user to speak inside the Hound App



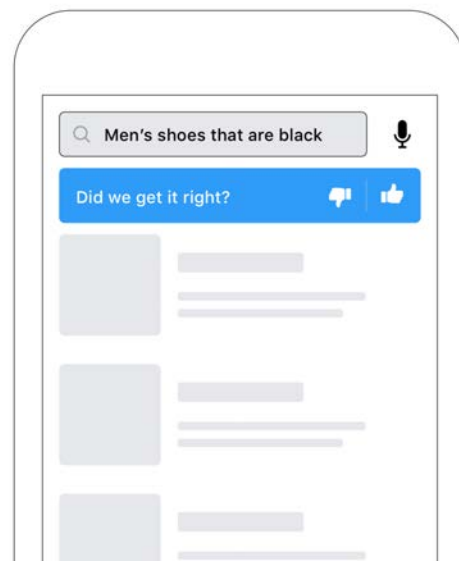
One solution is to simply tell the user exactly what is happening or what they should do. Use phrases like “Listening” or “Speak now” to help the user understand that it's their time to shine. Or if you'd like the user to feel as if they are communicating with a more “intelligent” entity, you can use first-person phrases like, “I'm listening” or “How can I help you?” These phrases can be shown immediately when the listening screen appears, or you can wait for a short time. By adding a slight delay, you avoid distracting users who know exactly what they are going to say.

## 4 Transcribing to show active listening

When the user begins to speak, a transcription is displayed so they can confirm the accuracy of their input. It's important to note that with our Houndify technology, transcriptions will change and update as the user speaks. This is because we gain context clues on the fly that allow us to more accurately understand what's being said, as it's being said.

Transcriptions end shortly after the microphone stops hearing voice input. A small pause (about 2 seconds) before completing the transcription helps to avoid cutting of a user's input too soon.

After the transcription ends, the text is sent to the server so that a response can be provided. We recommend using a sound effect here to communicate that transcription has ended and the search has begun.



The time it takes to get a response depends on several factors, like connection speed or complexity of the query. During this process, use some type of loading indicator to visually communicate that the search is being performed. Make sure to keep the transcription visible during the searching process so the user doesn't lose context (a surprising amount of users mentioned this in our user research). Once the search is complete, the listening screen will animate out and results should be displayed.

- **Call to Action:** Microphone button: some affordance of voice interaction must always be visible on screens where voice interaction is enabled — the most widely accepted form of visual affordance is the microphone icon.
- **Wake Word:** The existence of a wake word is not something users can see, so it must be taught.
- **Listening Screen:** The listening screen contains several key elements that help bring the experience together. A text prompt guides the user into action, and the audio input visualization (usually animated sound waves) communicates that the app is actually hearing what the user is saying.
- **Processing:** Use a sound effect to communicate that transcription has ended and the search has begun. Use some type of loading indicator to visually communicate that the search is being performed. Make sure to keep the transcription visible during the searching process so the user doesn't lose context
- **User Feedback:** Allowing users to give feedback helps you decipher similar-sounding terms. Secondly, a feedback mechanism allows users to feel empowered and provides an actionable follow-up to an incorrect query.

This feedback mechanism can be as simple as a thumbs-up or thumbs-down selector, or you can let the user type a more detailed message to explain what went wrong.

In the next chapter, learn how to improve user onboarding, education, and discovery. With voice technology, we have to learn how to talk to inanimate objects. Teach users a new way to communicate.

---

Chapter 05:

# User Onboarding, Education and Discovery



# Guiding Audiences Down the Voice-Enabled Path

*"A keyboard. How quaint."*

- Lt. Commander Montgomery Scott,  
Star Trek IV

In a new world of voice experiences, one of the biggest challenges is not so much the technology as it is convincing generations of keyboard users to keep using voice after they've tried it. After all, we've all been embarrassed by voice-to-text errors when we send text messages. That experience has made us a bit mistrustful of whether voice apps are up to the job. However, the new generation of voice AI is powered by machine learning and the quality of voice interactions is markedly better. Now, it becomes the job of the VUI designer to help increase the adoption of voice solutions through education, guidance, and a sharp eye on metrics.

Another challenge designers might face is timing. This is particularly true for products that haven't previously required or provided voice experiences. If a user opens your app and is immediately asked to provide microphone access, they may be thrown off.

While allowing access may worry some users, it's a necessary step for a successful experience.

When it comes to actions, voice interfaces are often billed as "intuitive." In theory, yes; but voice apps are not Jarvis or C-3PO, intelligent assistants who can understand any command; they're more limited than that. And yet—the more human an interaction (and conversation is among the most), the more likely people's expectations inflate. How do designers let users know all that is possible without overwhelming the user? And how do they prepare for everything a user might say? How do companies like Amazon and Google make discovering and enabling their voice apps easier?

When it comes to rewards, what kind of incentives will bring users back? Is the experience immersive, with high production values? What kind of voice products make the most sense on an (often) audio-only platform? And what kind of investments can users make on a voice-only platform? What can "lock users in," in a positive sense?

# Best practices:

## 1 Take every opportunity to provide education and guidance

The more people know about the capabilities of VUI, the more useful they'll find it. Many people tiptoe into the world of voice by trying several apps and then settling on just a couple that they figure out on their own. That's why it's up to designers and writers to help users find and remember the full potential of voice interfaces and to guide them through a discovery process. The type of guidance voice users require is more than a typical six-to-ten week drip email campaign. It's continuing education that highlights the features and capabilities of the VUI, reinforces what has already been introduced, and keeps users apprised of new functionalities as they come online.

A single source of education isn't nearly enough. It has to come from everywhere. Some ways to continually educate users include:

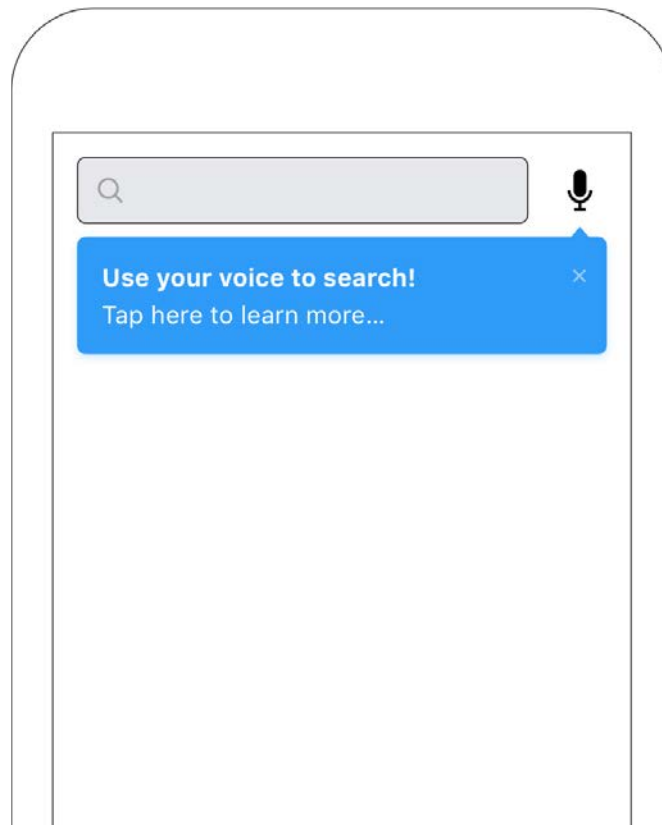
**Offer helpful tips:** One way of calling out this new experience is through the use of tips. A tip allows you to both educate the user, as well as specifically point out where the microphone button lives. At Soundhound Inc., senior product designer Erik Bue builds these tips into the interface. "We provide tips with every session," he said. "By providing different tips and hints each time someone visits the app, we help people discover new things that voice can help them with. Simple use cases like setting alarms and timers are pretty well known, but some of the more rich types of experiences aren't as obvious to people and should be pointed out."

Tips can appear within an app, as a voice message, in email, in text messages, or any number of places. Keep in mind that a tip should be actionable—something the user can try right away.

**"By providing different tips and hints each time someone visits the app, we help people discover new things that voice can help them with"**

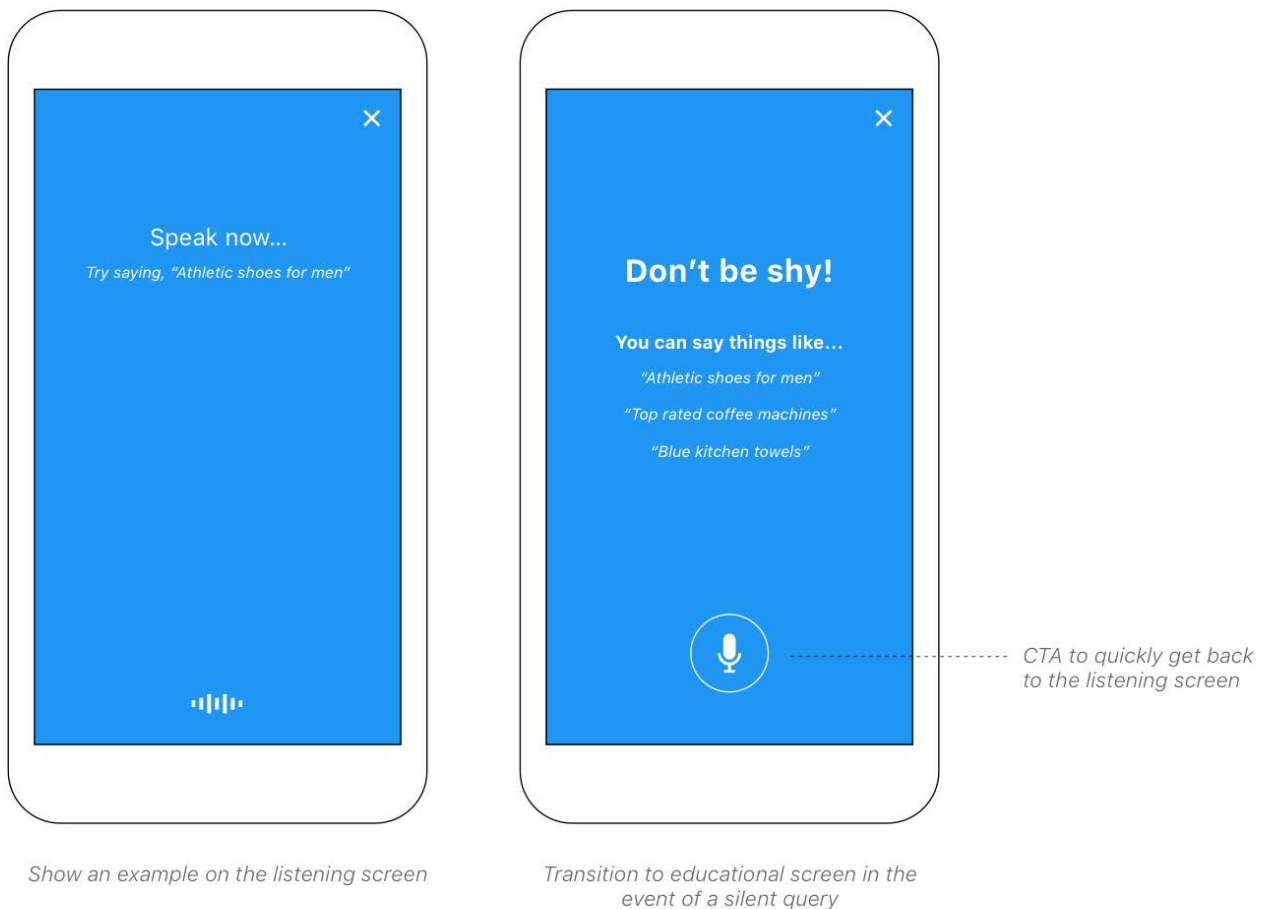


**Erik Bue**  
Senior Product Designer  
SoundHound Inc.  
🌐 <http://erikbue.com>



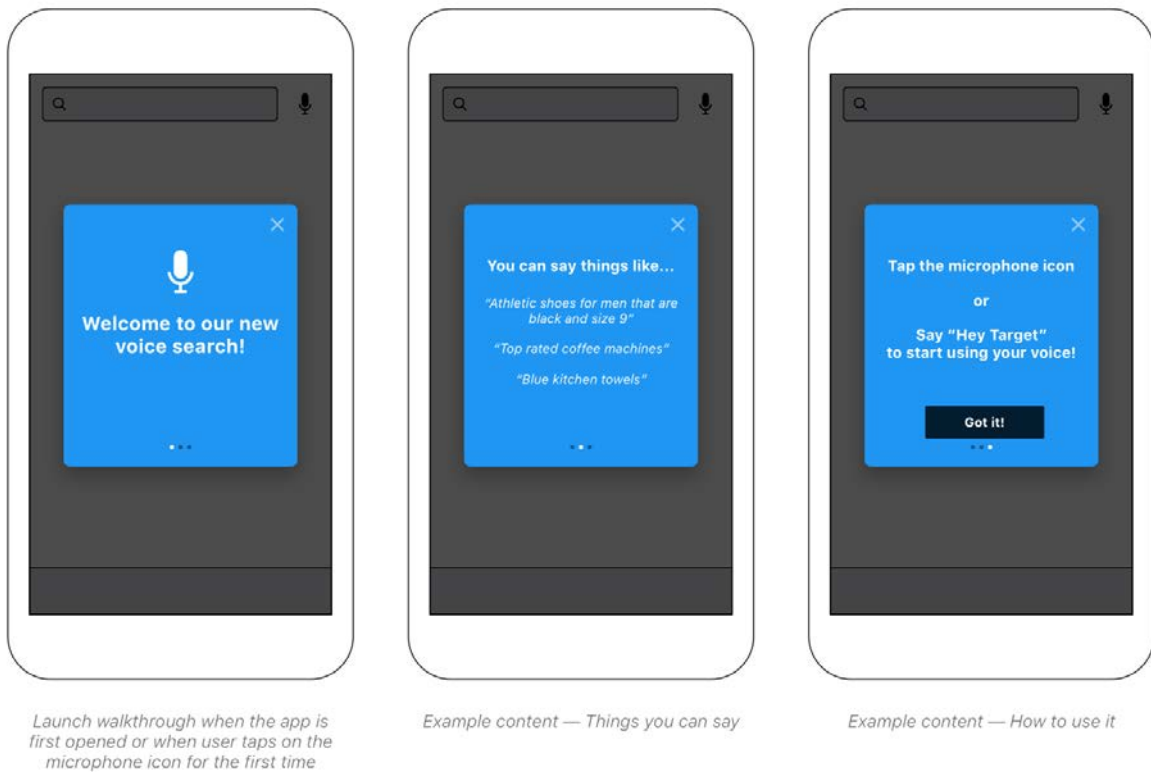
**Give them examples:** Another way to help users understand what to do during the listening experience is to provide specific examples of things they can say. Use simple and generic examples that could apply to a wide range of users. These example queries could be something like, "Black shoes for men" or "laundry detergent." If possible, you could also utilize previous search terms from the user's history to create a quicker and more personal connection. So, if the user has searched for "whitening toothpaste," you might use this as an example. We've found that showing one example at a time is most effective so the user isn't distracted by multiple items.

Additionally, if a user still hasn't said anything, resulting in a silent query, you can transition to an educational screen with more instructions or other examples of things they can do. This will allow them to pause from the listening experience and take a few moments to familiarize themselves. Remember to provide a call-to-action that will allow them to quickly get back into the listening screen when they're ready.



**Onboarding walkthrough:** If you want to be a bit more bold in introducing this new voice feature, you may want to use a walkthrough-type of experience in which you block the UI with a modal when the user opens the app. You can also wait until the user taps the microphone button or clicks on a tooltip. By showing it when the app is opened you guarantee that all users will at least have the chance of seeing it, but you risk annoying them by blocking them from their normal routine. When the user dismisses the walkthrough, you can use a tooltip to show exactly where to go to try what they just learned.

On the flipside, if you wait to show the modal until the user taps the CTA, you avoid blocking the UI, but you may not reach users that don't notice the button or don't feel comfortable trying something new. Depending on the amount of content you want to provide, this walkthrough flow can contain multiple pages (but try not to overwhelm the user).



**Teach a little at a time:** VUI and UX designer [Bryan Sebesta](#) teaches voice experience design at Utah Valley University, and he recommends educating users in small chunks so they don't get overwhelmed and give up. "Don't release new features all at once, or at least don't advertise them all at once. We're limited by our memory, which can only take in so much at once. Introduce things one at a time, and then remind users of them often, so that the features and abilities make it into long-term memory."

***"Don't release new features all at once, or at least don't advertise them all at once. We're limited by our memory, which can only take in so much at once."***

"The content itself is the educational piece," says Erik Bue, senior product designer at Soundhound Inc. "Within certain interactions, if I ask, 'Show me



restaurants nearby," we also provide suggested follow ups to that question. For example, "You can ask which restaurants have free Wi-Fi." These little hints help you understand how to ask questions better the next time."

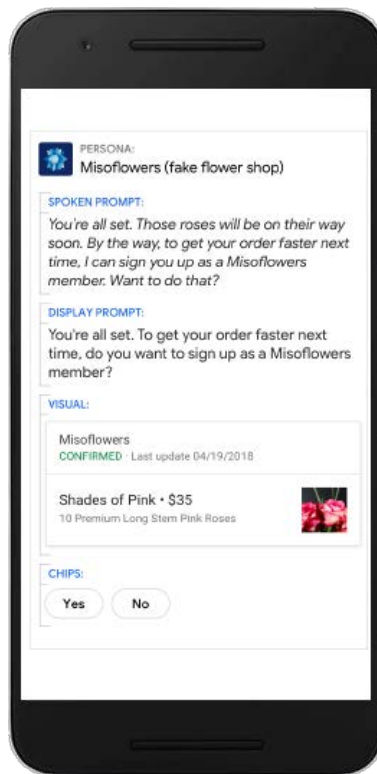


Amazon effectively advertises on their website with small blue bubbles containing italic text, that Amazon's customers have come to associate with Alexa speech bubbles. Help users understand what to do during the listening experience by providing specific examples of things they can say.

**Cooperate with the user and their memory:** A top priority for voice design should be to cooperate with the user. For our purposes, this means: be brief, don't overwhelm people, and work with the constraints of people's working memory. When you first open a skill, [give a few suggestions](#). As the user continues to use the action or skill, you can [taper](#), meaning you provide fewer cues (or provide different cues, suggesting different features). Other strategies, like [providing overviews](#) and keeping any list of options to fewer than four, can also be helpful.

**Prompt users to ask or try other things:**

Knowing what to say without overwhelming the user is an art, not a science. The goal is always to cooperate, and that includes not underwhelming the user, either. With voice, this often means providing clear overviews of how much information we're about to present, and how it's structured, being concise, and knowing when to break up a lot of data into several turns, and what's most relevant. For example, an airline might present flight options by listing out four times. When a user selects the time that works best for them, only then might we explain what time works best.



In this example, the voice assistant suggests new options within the conversation, but presents it in small, manageable pieces.

## 2 Data is key to improving voice experiences

Heidi Culbertson, founder and CEO of [Marvee](#), a company dedicated to voice design for older adults, cannot stress the value enough of collecting user data and metrics. In fact, it's paramount to providing the best user experience and increasing adoption and retention. "Voice AI is only going to improve based on the data that we can feed into it," she explained. "And a lot of that comes from user data."

[Pandora](#)'s Ananya Sharan, a product manager for the music streaming service's voice mode, expanded on the idea. "The more a user engages with a VUI, the greater the volume of data you can get. When we learn more about what your listening habits tend to be, thanks to machine learning and AI, these can all be automated to deliver the right result to you."

How should you use customer data? Conduct ongoing A/B tests to find out what your users prefer. Then, build a programmatic and systematic roadmap to increase user engagement. Through behavior, users will tell you what's working for them. "So many mistakes can be avoided by talking out loud with

other people early on," Sebesta said. "This helps you catch the rhythm of conversation at the first stage, and keeps you from "writing for screens," which is what most of us are trained to do. It also keeps you from saying too much at once. Talk out loud several times before committing anything to type."

**"The more a user engages with a VUI, the greater the volume of data you can get. When we learn more about what your listening habits tend to be, thanks to machine learning and AI, these can all be automated to deliver the right result to you."**



**Ananya Sharan**  
Product Manager  
Pandora - Voice Mode  
🐦 @\_ananya\_

Measuring success isn't purely determined by volume when it comes to the addition or creation of a VUI. Some of the metrics you'll need to monitor include:

- The percentage of voice users who use voice functionality repeatedly
- How and when users take advantage of voice capabilities
- Overall satisfaction with the app from voice users
- The degree to which voice users feel the overall app experience is better because of voice

### **3 Deliver real value**

When it comes right down to it, what users want from their voice assistant is value. Make something that makes using a VUI worth it. It doesn't mean what you make has to be complex or involved. The ability to ask "when is my next appointment?" and hear a short response can be exactly what a user needs. Voice apps do not need to have a massive range of use cases and abilities. Often, that can only make every feature harder to remember.

***"Voice has to provide something that cannot be accomplished as easily if you were typing and swiping."***

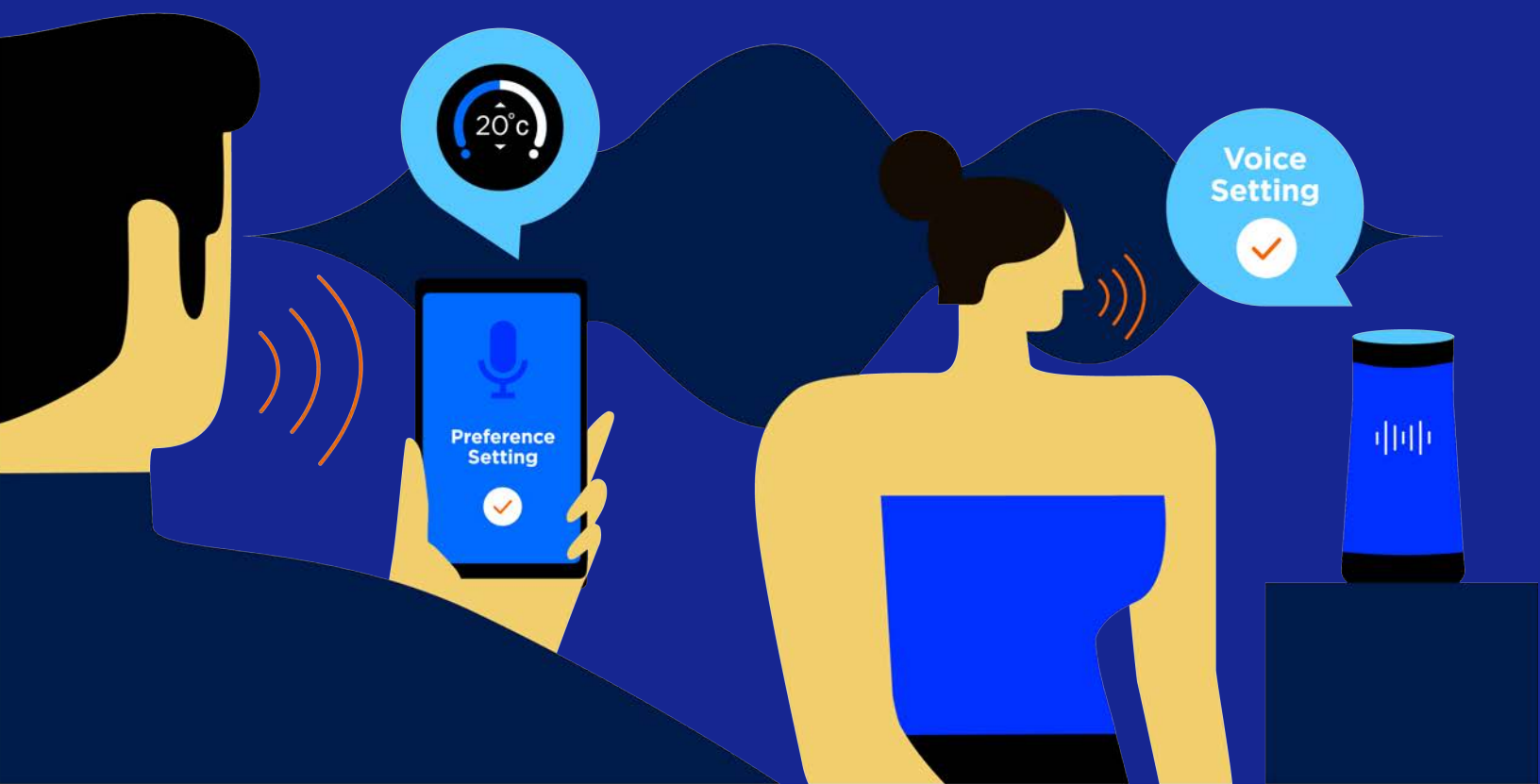
Sharan has other ways to measure value. Her customers could easily search for the music they want through the Pandora app, so she must deliver a compelling reason to use voice instead. "What is the incentive for the listener to use voice instead of going and tapping it on the app?" she asks. "Making it effortless and easy drives adoption. And keeping it easy means we have to get it right every time." Not only that, but voice has to provide something that cannot be accomplished as easily if you were typing and swiping. "If I were to go into the app and search for 'something relaxing,' I could go into the browse modules, look at the chill stations, look at nature sounds, and choose something. But I'm doing all the work. With voice, I don't have to do all that. I can just say 'Oh, I like this song,' and it's automatically added to my playlist, and I'm done."

In the next chapter, learn how to enable users to choose their own experience by adding some personalization. As we build rapport with our VUI, it should remember our preferences. Let the learning continue.

---

Chapter 06:

# Personalizing User Experiences



# A Voice That's Uniquely Yours: VUI Customization Options

We've become a society obsessed with personalization. It's all the rage to tailor everything from hamburgers to mattresses to cars to even the brightness and color of the lighting in our environment to our personal preferences. It's not enough anymore for us to just have personalized shoes, food deliveries, and playlists. We now expect brands to know about our preferences in advance. Technology has, in that way spoiled us. Just look at our personalized social media pages, we see ads for items we're already shopping for and events we would definitely be interested in attending. We can watch Netflix or YouTube which analyze our viewing patterns and predict how much we'd like a different show—and it's almost always right.

Should you give users options to personalize and customize interactions based on their preferences and previous interactions? How does this help increase utility and strengthen the bond with the user?

First and foremost, personalization should be seen as a way to provide a better user experience. Just adding a name to an email blast isn't going to cut it. And that's where the challenge lies. At the heart of personalized experiences is user data—and a lot of it. To successfully create a personalized experience, we need to be aware of the different user journeys and be smart about how we analyze and use data.

***First and foremost, personalization should be seen as a way to provide a better user experience.***

Fortunately for the burgeoning VUI industry, this kind of personalization lends itself well to voice applications. Leveraging voice capabilities to provide a more customized experience depends almost completely on better design and data.

# Best practices:

## 1 Adapt the interface to the user's personal context

If a user always asks for the same series of things, we can “remember” it and put it forward in the next interaction. “This works much better in voice interfaces than visual ones,” according to [Paul Zumbrink](#), director of user experience for voice and conversational AI agency [RAIN](#). “Change and adaptation are part of our natural way of communicating.”

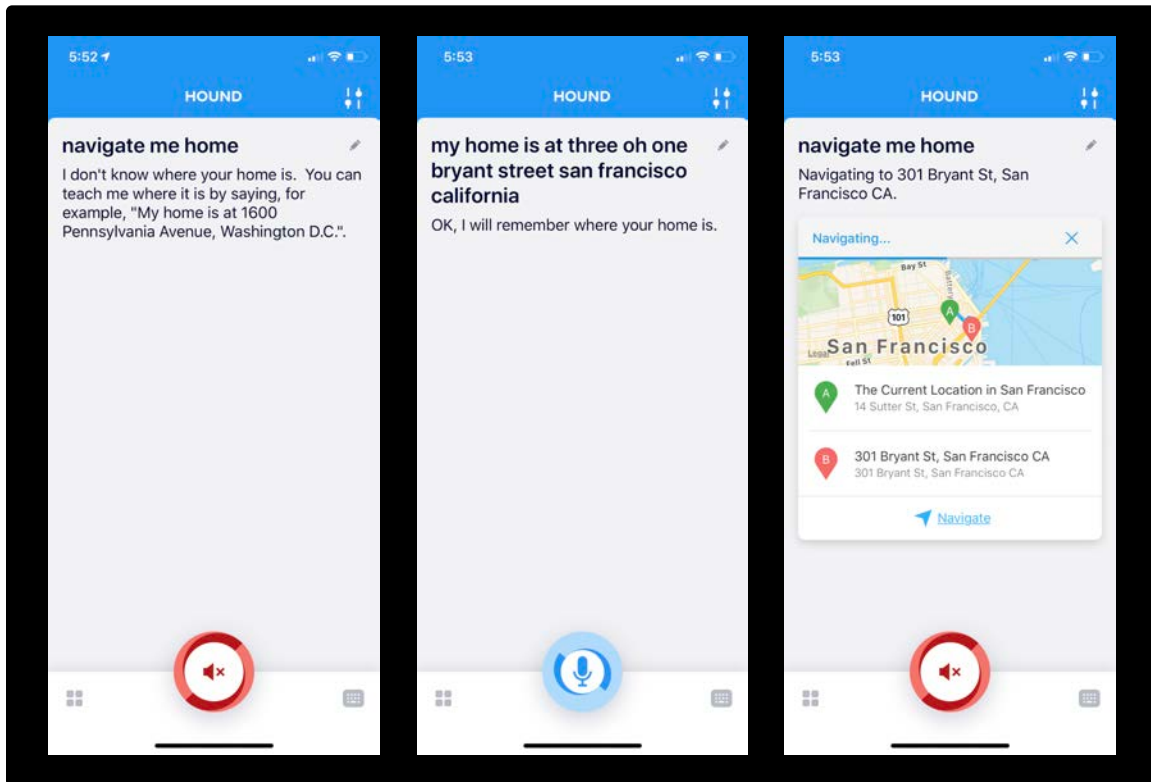
Predefine a classification system for different interaction metrics, Zumbrink advises. The taxonomy can range between some of the following (depending on the product):

- The current user's actions
- The previous user's actions
- The user's physical context (location, time, day, weather, etc.)
- The user's system context (number of sessions, average session length, etc.)

“It's possible to further classify each user in a different tiered group, such as beginner, intermediate, or advanced, and then have the interface gradually shift based on how many times the user interacts with the app, and how accelerated their journey was,” Zumbrink said. With your users classified into personas, it's possible to personalize everyone's experience based on their usage and knowledge and “grow” along with them.

Another consideration around personal context is apparent with use cases like Pandora, where personalization is “absolutely number one for us,” according to Ananya Sharan, product manager for Pandora's voice mode. Customizing the music experience is at the top of their list. As we learned in previous chapters, carrying your brand message and values through the VUI's personality and functionality is a key consideration, and Pandora is a perfect example of why and how to do it right. Their entire business differentiator is based on knowing what their customers like and want in advance.

With the Hound App, users can personalize multiple things like saving their home and work addresses — so the user can say “Navigate home,” or “What’s the weather like at home today?”



## 2 Think about the wider system where a product will exist

Personalization is not the sole answer to successful products. To create the most compelling experience possible, imagine your product “in the wild,” and how it can be most useful to each user. Zumbrink recalls creating the voice assistant for Starbucks. To allow for the best experience, he considered how it might be used in a busy urban environment like Manhattan.

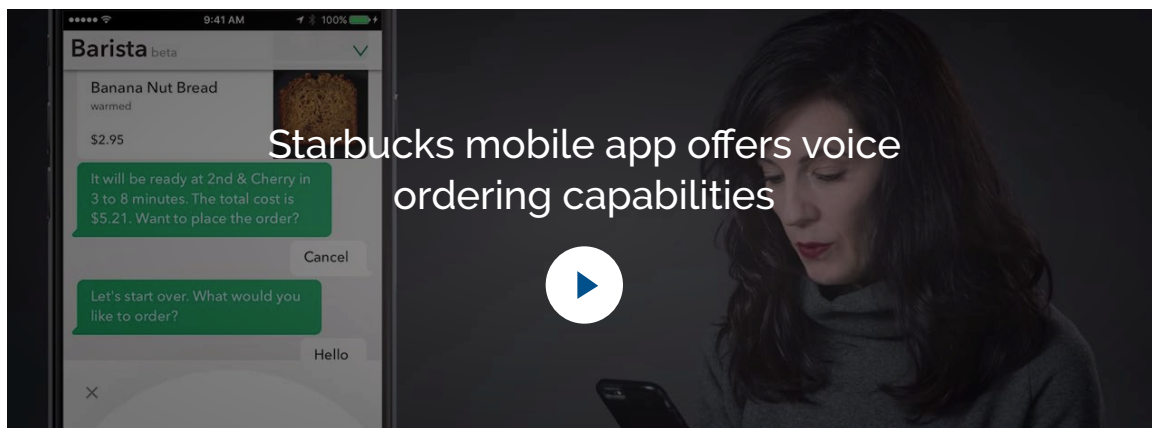
For the Starbucks voice assistant, they not only leveraged “your usual drink,” but also:

- The location of the user and direction in relation to the closest shop
- The ecosystem of other Starbucks apps and tech
- The primary Starbucks card balance and reward points
- The different users who might have access to the app in the household



"So if I ask for my 'usual drink,'" the following will happen'

- The app will know it's me, and that my usual drink is a triple grande macchiato in a venti cup
- By the time the drink is ready, I'll be closest to the Times Square shop
- When I get to Times Square, I'll get a notification on my Apple Watch that my drink is ready
- In the shop, I'll pay for this entirely with my reward points"



The idea, as with Pandora, is to make the entire experience easy and frictionless, and with as little work from the user as possible. Behind it all is the data you need to collect, and the metrics you've captured around the different flows that people take when conversing with the product's voice assistant.

Mercedes brings this to life in their cars. Mihai Antonescu, a senior engineer at Mercedes-Benz R&D explains: "The first time you use MBUX you'll be creating a profile for yourself as the driver. And then the learning experience starts for the system. So depending on where you're going, who you're calling, what you're listening to, which radio station you play and so on, the system will learn that. So the next time that driver's in the car, it will recommend those choices. You're just one click away from achieving that action."

"Depending on where you're going, who you're calling, what radio station you're listening to in the car, the system will learn that. So the next time that driver's in the car, it will recommend those choices."



**Mihai Antonescu**  
Senior Engineer  
Mercedes-Benz R&D  
North America  
🐦 @MBRDNA

### 3 Tag and track interactions and visualize user behavior flows

"At RAIN we use a proprietary technology platform [VOXA](#) to tag and track interactions, which we visualize through analytics platforms such as Dashbot and Google Analytics," Zumbrink said, adding, "but the most common use of personalizations has to do with recommendations based on plug-and-play machine learning algorithms, like [Amazon Personalize](#)." This is where you keep track of specific user preferences and use this to provide the user with other product recommendations."

Inserting product data and metrics into a product analytics dashboard will greatly help to visualize user behavior flows and identify drop off points. "In some cases," Zumbrink said, "we learned a great deal just from observing user flows that proved or disproved our hypotheses."

If you haven't read all six chapters of this VUI best practice guide you can find the full list on the [guide's homepage](#).

