voicebot.ai™
GIVING VOICE TO A REVOLUTION

APPLAUSE

VOICE UX BEST PRACTICES

LESSONS FROM 17 EXPERTS

Table of Contents

About Voicebot

Voicebot produces the leading online publication, newsletter and podcast focused on the voice and AI industries. Thousands of entrepreneurs, developers, investors, analysts and other industry leaders look to Voicebot each week for the latest news, data, analysis and insights defining the trajectory of the next great computing platform. At Voicebot, we give voice to a revolution.

voicebot.ai™
GIVING VOICE TO A REVOLUTION

About Applause

Applause is the worldwide leader in crowd-sourced digital quality testing. With 300,000+ testers available on-demand around the globe, Applause provides brands with a full suite of testing and feedback capabilities. This approach drastically improves testing coverage, eliminates the limitations of offshoring and traditional QA labs, and speeds time-to-market for voice, websites, mobile apps, IoT, and in-store experiences.

Thousands of leading companies — including Ford, Fox, Google, and Dow Jones — rely on Applause as a best practice to deliver high quality digital experiences that customers love. Learn more at **www.applause.com**.

APPLAUSE

PEOPLE ARE TALKING

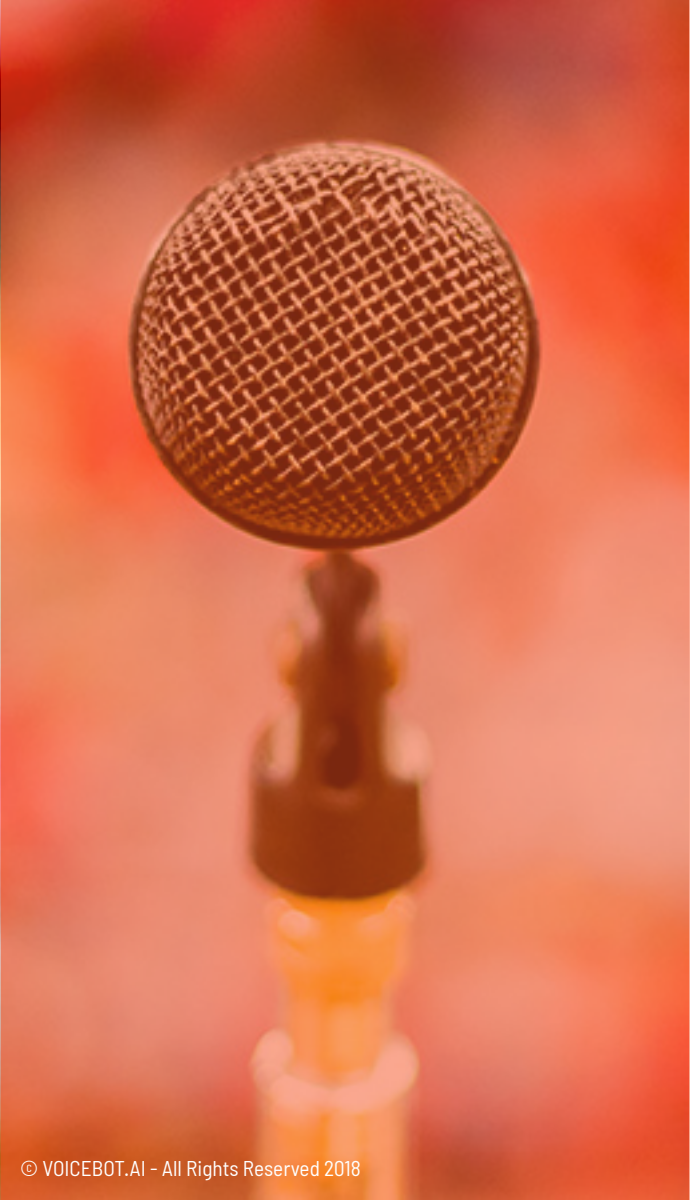
Voice is changing the way we interact with our personal devices—even within our larger environment. From telling Alexa to set a rice timer for 10 minutes to asking Google Assistant to make calls on your behalf, interactions with voice AI are becoming increasingly common as technological capabilities evolve. Voice isn't exactly new even in the technology space. Speech recognition has been quite capable for at least 20 years and Siri has been helping us send text messages since 2011.

What is different is the rapid expansion of use cases and access to voice assistants. Whereas voice technology was once synonymous with the dreaded interactive voice response (IVR) phone trees that were designed to help curtail customer service costs while wasting time for consumers, voice assistants have changed perceptions and accelerated frequency of use seemingly overnight. Voice is everywhere. But, are we implementing these new solutions properly?

Teaching Computers to Understand Us

Using voice to access digital content and execute tasks is very different than click, touch and swipe. Previously, we had to learn how to speak a computer's language by either employing manual inputs or occasionally adhering to a narrow set of speech interactions that digital devices were trained to understand. We aren't learning how to talk to computers anymore. We are teaching computers to talk to and understand us. The shift is profound. It places a larger burden on voice app developers to deliver a seamless voice user experience in order to lessen the burden on the user.

However, you are not alone in confronting these challenges. Other people have accumulated decades worth of expertise in designing voice user experiences while some have come to the discipline recently but are already offering many new insights. Over the past year, the Voicebot Podcast has conducted over fifty hours of interviews and many of those discussions centered on voice user experience design. Some of the guests include authors and long-time voice UX design practitioners such as Cathy Pearl and Lisa Falkson. Other guests such as Tim McElreath and Mark Webster brought visual design expertise to the problem and learned how to modify it for voice. These designers by training have been complemented by linguists and technologists such as Tobias Geobel, Noelle LaChartie and Jan König that introduced an entirely new perspective.

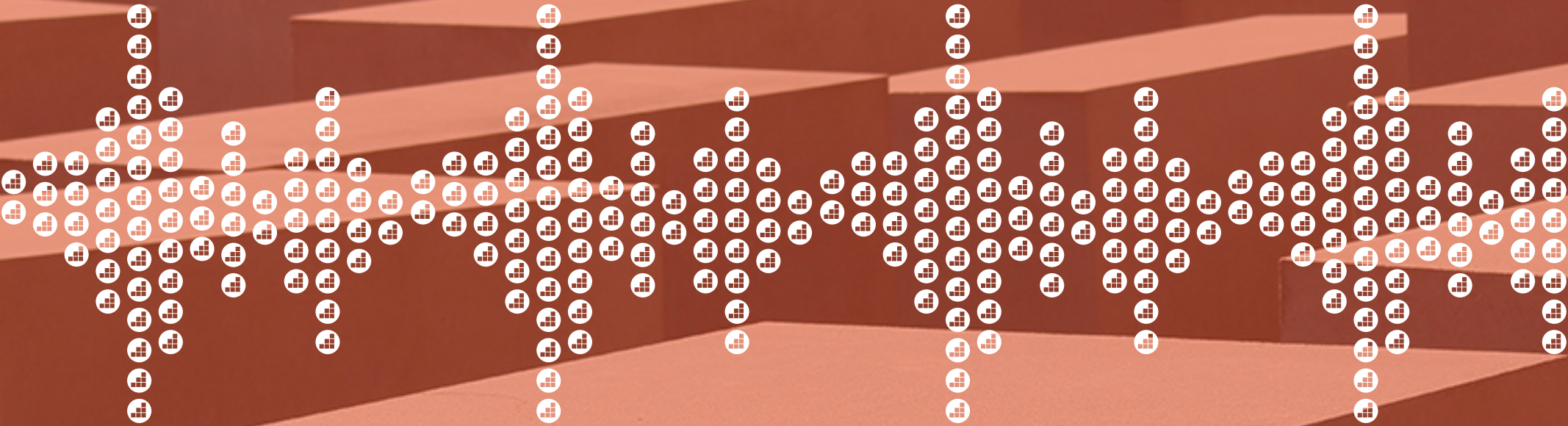


In all, we have amassed over 100 insights, tips and techniques from 17 voice UX experts in the Voice UX Best Practices. The eBook was created for designers, technologists, marketers and managers, especially those focused today on building for Amazon Alexa, Google Assistant and other emerging voice assistants. It provides a quick overview of the most common challenges, pitfalls and opportunities associated with voice UX. And, it will hopefully become a trusted resource to share with your teams when they need to get a quick overview of key considerations associated with the discipline. The eBook is organized in four sections with 12 subsections listed on the right side.

We hope everyone finds the information helpful and easily accessible for the experienced practitioner and novice alike. There will be a companion podcast where you can hear from the experts themselves in their own voices as a supplement to these materials, but both will stand on their own. If you have questions about the content or topics that were not covered, feel free to connect with Voicebot on Twitter or email us at info@voicebot.ai.

EBOOK SECTIONS

FOUNDATIONAL CONCEPTS



"It's not the user that is the problem. It's not just that they have to learn the software or the program...it's that we as designers should do a better job modeling it around the way humans are."

CATHY PEARL
Author, Designing Voice User Interfaces

Building for Users

You aren't just building a voice experience—you are building for an audience. People are on the other end of the AI and will use and interact with what you create. You are designing for them. Therefore, it is crucial to keep your users in mind during the entirety of the build and test process, so that the user experience can ultimately be as seamless and as simple to navigate as possible. The user experience is the core responsibility of the designers, but everyone has a role to play. Developers, writers, quality assurance and everyone else that works on a voice app from concept through ongoing use and refresh have an impact. The number one rule should be: "Don't blame the user. Instead, help the user. Make it easy, efficient and delightful."

In the early days of computing technology, we had to write in the computer's language—punch cards, binary code. Now computers are integrated into our day-to-day lives and we are teaching them to communicate in our language. This means that computing is increasingly more accessible. As designers and developers, it is our responsibility to make computing accessible and worthwhile for users—not so that it takes up more time, but ultimately gives them time back by making computing tasks easier.

"As we shift towards a more casual computing context, where people are dipping in and out of computing environments and expecting a lot of personalization and customization, there is an opportunity now to sort of shift the dynamic. The hope and the idea is to give a little more control back to individuals and to make the technology more responsive to individual's needs.

CHRIS MESSINA
Product Designer

Language is the bridge of communication between humans and computers. It is important to clearly relay to users how they can communicate with computing devices to receive the information or interactions they desire. Set expectations. Don't overpromise. Be realistic. Robots and assistants, while built to be increasingly human-like, are not human, and have different capabilities.

> *"We should design our robots and assistants to be human enough but not too human lest they overpromise. And of course, we have to define what 'human enough' is…We don't want to trick people into thinking that our robots are human because we're just going to let people down. So setting expectations of what the robot can do becomes huge," Karen Kaushansky, Consultant, Robot Futures Consulting*

Another critical aspect is designing with both the user and the use case in mind. We all know that groups of users are not monolithic. However, keep in mind that an individual user has a variety of expectations depending on context and use case. If the use case is transactional such as sending a text message, the user might highly value efficiency. In another transactional use cases, comprehensiveness might be most important. If the use case is immersive such as entertainment or meditation, the user may value efficiency and comprehensiveness less than fun, relaxation or learning. And, consider the variety of emotions and expectations which are often associated with these different use cases. These characteristics can make or break the UX.

> *"Make sure your tone matches your use case. Take into account the emotions people are going to be bringing into your interaction, and be very clear about what type of emotions you want people to carry away from your interaction," Tim McElreath, Director of Technology, Discovery Communications*

To Build Better For Users

1. Model your programs around the way humans are and expect to interact.
2. Be clear about the capabilities of your app; don't over-promise.
3. Acknowledge your users.
4. Make sure the use case guides your decisions

EXPERT EXAMPLE
CATHY PEARL

"If you want a date from somebody—you're doing a travel app—and you say, "when do you want to travel?" You might have all your dates plugged in, and then someone says, "I want to leave next Tuesday evening." Maybe you did not expect that. Modifying prompts is a simple solution: if you prime someone by saying "what date do you want to leave?" they are much more likely to give you an actual date."

Building for Conversation

When building for conversation, don't assume you know how people talk. Not everyone relays information and conveys meaning in the same way. Even when you think a particular response to a prompt is obvious, that is rarely the case. This is the crux of how conversation is a fundamentally different way to interact with technology than where we started.

> *"Design for how people actually talk, not how you want them to talk. You can modify prompts to cater to this, as the prompts themselves are key in getting back the kind of information you want," Cathy Pearl, Author of Designing Voice User Interfaces*

Accommodating the Computer's Limitations

Historically, the communication and understanding capabilities of computers was severely limited. Computers only knew their own language and input mechanisms were devised to enable humans to input instructions. Initially that meant communicating directly in the computer's language by typing out code or commands (or using punch cards created on special typewriters) on a screen.

GET CONVERSATIONAL

TIM MCELREATH, Discovery Communications

"Don't expect that people talk the way you talk, or converse the way you converse. It is crucial to assume that there's going to be a diversity in the way people converse even within a single language."

"One of the things I learned very quickly is in order to design an effective voice experience you have to listen to the response a lot, over a long period of time. You don't know how frustrating a response can get until you've heard it 50, 100 times."

"If you want to design an experience that people start using visually, and come back to over and over again, you have to design it so that they are going to want to sit through a response to the end without getting frustrated."

This interaction method was displaced with the introduction of new user interfaces (UI) that abstracted the communication method and translated more human-friendly input methods into computer language in the background. In turn, these UIs translated the computer response into formats more easily consumed by humans.

However, these were all programmatic interfaces. There were a predefined set of interactions with no way to go beyond them. Menus and buttons made it easier to interact with computers while clearly proscribing the limitations of our interactions. If the option was not on the screen (i.e. the UI) then it wasn't an option. This reduced the incidence of errors, but forced users to learn the systems in order to use them. Humans by necessity adjusted their communication methods to account for the limitations of computers.

Beyond Programmatic Interfaces

The advent and advancement of Natural Language Processing (NLP) with its Automated Speech Recognition (ASR) and Natural Language Understanding (NLU) capabilities introduced the option for non-programmatic inputs. It has flipped the dynamic enabling humans to communicate in their preferred method and computers to accommodate, adjust and learn. It's an extraordinary change and UX designers are vested with the responsibility to help computers succeed in the new dynamic.

This is the context behind the UX designers requirement to design for conversation. It is not just about helping the user, but also helping the computer, the voice assistant, be successful in its interactions. So, designing for conversation presents two challenges. First, you must design for both the user and computer. Second, you must design for interactions without the boundaries many of us became accustomed to in programmatic interfaces. As you move into conversational design, experts say critical success criteria include not assuming you know how users will speak and evaluating actual conversations to really understand what is working and what is not.

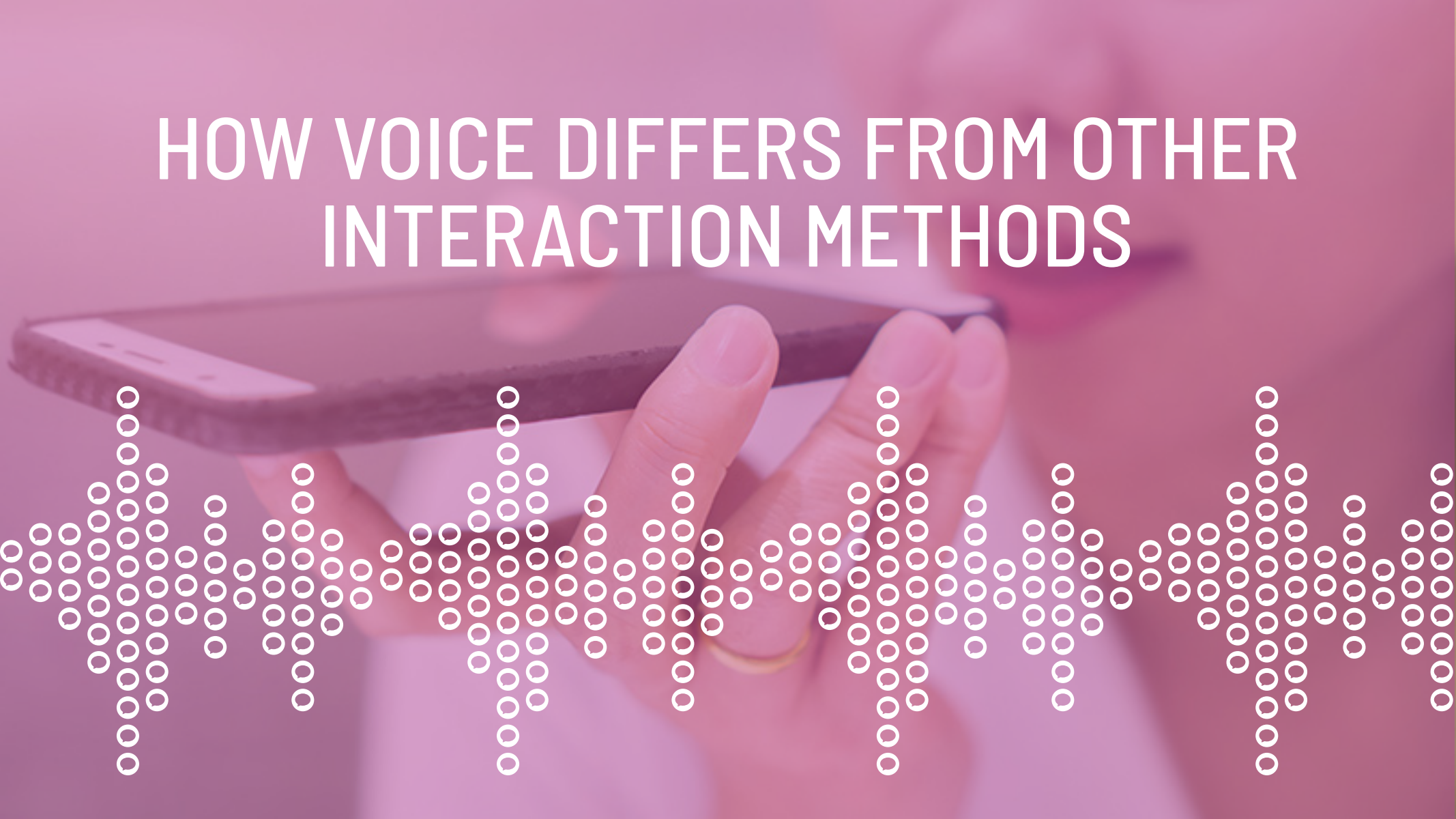
To Build for Better Conversation

1. Design for how people actually talk. Assume diversity in the ways people talk and converse.

2. Use carefully-constructed prompts to get back the information you want.

3. Don't bury your prompt in the middle of your response. Put the prompt at the end so it is the last item the user is reacting to.

HOW VOICE DIFFERS FROM OTHER INTERACTION METHODS



How Voice Conversation is Different from IVR

The perception of many consumers about voice interaction with computers was initially shaped by interactive voice response (IVR) systems implemented by banks, government agencies and just about every other bureaucracy that must field customer inquiries. In turn, the methods employed for these use cases also shaped the biases of many voice user experience designers, for better and worse. Speech recognition performance and the application of natural language understanding (NLU) has moved well past the early days of IVR.

> *"Now I don't have to write these careful grammars. I can accomplish more in a single prompt than I could before. So, I have to remember to open it up again, [to not constrain the experience,] because you can do a lot more successfully [since the systems have greater capabilities]," Cathy Pearl, Author of Designing Voice User Interaces*

Voice Recognition is Not Inherently Conversational

It is important to distinguish between voice recognition based inputs and conversation. IVR systems were not, are not, conversational. Most are programmatic interfaces that substitute a spoken word or short phrase for a mechanical input. It's like a voice-click. You can travel down the phone tree, but there are no alternate paths. The nuance and freedom provided by spoken language is unwelcome. Very often only the choices included in the prompt are accepted and they are utterly void of meaning.

Conversational interaction is different. The entire focus is on understanding intent and determining meaning to decide the proper response. A conversational interaction may offer a prompt to the user, but by definition should be capable of accepting responses outside of the expected conversational flow.

> *"We're all so used to interacting with voice interfaces through IVR systems on telephones...So, I call American Airlines, and say things like "save reservation, save booking," and we see a lot of people basically bring that same approach to Alexa, to Google Assistant...making it keyword-based, which is super problematic. It's easy to mishear one word. And, I as the user have to hold on to what all those available keywords are in my head, and [we are] not using enough of how you ask the question to influence how somebody responds to it," Mark Webster, CEO, Sayspring*

> *"Usually people are not calling in to talk to a voice automation. They're calling in to talk to a person and so it's always they call in to get this thing. They're like, I have to put up with it and so off the bat the conversation doesn't start on a good note."*

DR. AHMED BOUZID
CEO, Witlingo

IVR has had a lot of negative overhang from the past. Conversational voice technology is changing perceptions quickly as designers better understand how to design for conversational UI and users become exposed to conversational systems.

> *"One of the frustrations for me [when I was working with IVR] was certainly that so many times our clients said, "Look. We're trying to cut costs. This is why we built an IVR. We don't want people to go to the agents, so make them say 'operator' six times before they get to speak to an agent." But there are certain automated tasks people would be happy to do in voice and there are certain tasks for which they really want a person—so just separate those and just let them talk to a person. That was always kind of a battle," Cathy Pearl*

Voice Assistants Introduce a New Set of Use Cases and Expectations

Not only is the technology more capable, but the use cases and expectations are different. Dr. Ahmed Bouzid started working with voice technology in 1995 and has seen the evolution of voice technology first-hand from the early days of ASR to IVR to the recent crop of consumer voice assistants:

> *"If you were working in the voice space, you were really pretty much confined to what would make money. You were confined to delivering voice on the phone...meaning when you call into a system, you get to talk to that system."*
>
> *"Usually people are not calling in to talk to a voice automation. They're calling in to talk to a person and so it's always they call in to get this thing. They're like, I have to put up with it and so off the bat the conversation doesn't start on a good note. Whereas with Alexa and Google Assistant and so on, you intentionally or knowingly are talking to a robot or something that is not human. You don't perceive it as something that comes between you and a human which is what people perceive of these ideas. So, off the bat, the expectations are right, the disposition towards the technology is positive and it gives you a lot of leeway and a lot of goodwill from the user."*

THE PROBLEM WITH INTERACTIVE VOICE RESPONSE
TOBIAS GOEBEL, VP Product Marketing, Sparkcentral

"The speech-to-text part is where IVRs also struggle because over the telephone line a lot of the frequency range is cut off for bandwidth [reasons] still today. And, that is a problem obviously for speech recognizers. Now that is not so much a problem for an Amazon Echo which can benefit from the entire frequency range. So, you can have improved speech recognition models now and the speed of processors and things like that pretty much have solved this to the effect of working reasonably well on non telephony channels."

Voice UX practitioners working on rules based IVR systems learned a lot about designing voice interactions to accomplish a specific set of goals. These "jobs to be done" were two-sided in that the user had an objective to complete a task and the company wanted to reduce the number of humans required to service these tasks. However, the fallback was a human when the errors accumulated or the user insisted. Voice assistants are also designed to complete tasks, but they typically have no human fallback. That means users are generally more accommodating when encountering errors, but it also means designers are working without a safety net. They must account for many more variables, enable more paths to success and learn how to gracefully handle errors.

How to Avoid IVR Traps and Improve Voice UX

1. Let users talk like a person and do not force them into reciting a narrow set of rigidly defined prompts.

2. Don't constrain the experience. Think more broadly than the happy path you designed on the whiteboard.

3. Consider how your errors can generate graceful fails that help users get back on track because sometimes they will get stuck. Figure out what your fallback is when a human is not an option.

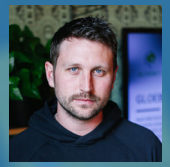
How Chatbots and Voice Are Different

Chatbots were all the rage before voice assistants started to experience rapid consumer adoption. As a result, many businesses experimented with chatbots before building voice apps. Chatbots offer a type of conversation. Voice assistants are inherently conversational so it seemed to follow that building voice apps would simply be about applying the existing chatbot design. However, chat and voice are as different as IVR systems and voice assistants.

Synchronous Versus Asynchronous Communication

One critical difference is that chat is asynchronous communication (or can be) whereas voice is synchronous. This has far-reaching implications from a conversational design perspective. First, in chat both parties generally have the entire conversation to refer to when formulating responses. In voice neither party is certain when the other is going to stop talking. It is not just about barging in and being able to interrupt. The uncertainty is also about when the microphone opens and when it closes. Are you talking to someone else in the room when the microphone opens and then turn your attention to the voice assistant? How does the voice assistant determine what speech was intended for it and what to ignore?

> *"One [difference] is synchronous and one is asynchronous and what I mean by that is if you talk to Google Assistant, it always talks back and then you have to respond. So, the design of the errors are very different. In the Google Assistant world or in the Alexa world, I don't know when [the speaker] is going to stop. In a messaging world, I get a message and I can decide when to respond. I can see the whole message. I know how long it is. So I can say something and I know what I'm going to input...You don't have to ignore anything. You look at everything. It's intentional. I meant to write this. [With voice] a lot of it is accidental. You have to design for what to ignore," Shane Mac, CEO, Assist*



CHATBOTS VS. VOICE

SHANE MAC, Assist

"There are major differences [between text chat and voice chat]. From a strategy level, what's interesting, is that every brand believes they are the same. So every brand believes that 'if it should work on messaging, it should work on voice' and vice versa. That is actually not true. With the right software the conversation on messaging and voice should live on one team but the thinking has to be different."

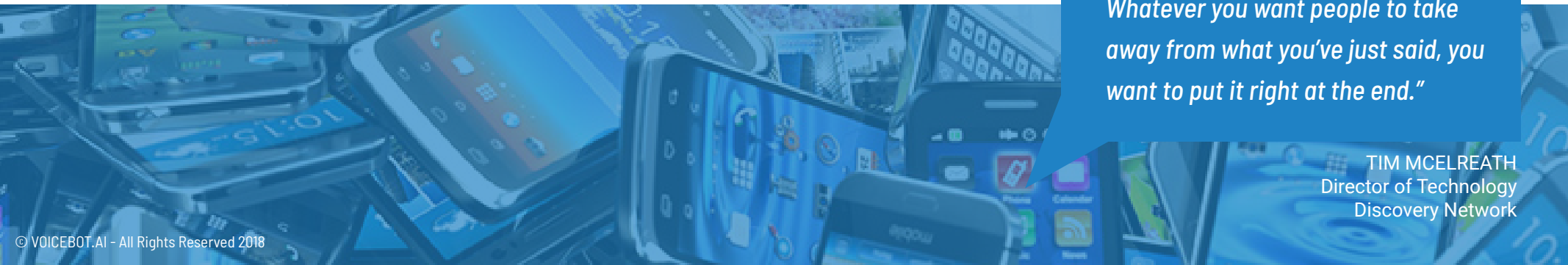
Similarly, what if the user pauses to collect their thoughts mid-way through speaking? The voice assistant microphone generally recognizes pauses, unless very short, as the end of the speech and then goes about intent extraction and formulating a response. There is good reason for this. The voice assistants want to appear to be as fast and responsive as humans. A long pause on the part of the voice assistant in an attempt to ensure it is allowing the user to complete their intended speech could be viewed as slow or unresponsive. This "feature" is not a bug, but invariably leads to errors in the conversational interaction. Chatbots don't have this constraint. As user experience designer Karen Kaushansky points out, because chatbots are asynchronous they allow users to proceed at their own pace or at least offer much greater flexibility.

> *"The power of conversational assistants and chatbots is letting users interact instantaneously to get what they want so it eliminates browsing, searching, clicking, and it allows them in some cases to do it at their own pace," Karen Kaushansky, Consultant, Robots Futures Consulting*

Chat Offers Visual Aid Where Voice Does Not

A second difference is the ability of the user to refer back to the transcript of the conversation when formulating a response. Think of this as text permanence or reference. The voice assistant may have this data to refer to, but the user typically does not because the voice assistant's audio response is ephemeral. Users only have what they can recall to help move the conversation forward. This means they may need information repeated in order to refer back to it or may forget about it entirely. In addition, techniques of effective communication differ for reading and listening according to Discovery Network's Tim McElreath.

> ***"If you're reading something off a screen, you want to put the major point [at the beginning]. The big takeaway should come first, because people will read the first sentence and [often] skim the rest. In voice, it's the exact opposite. Whatever you want people to take away from what you've just said, you want to put it right at the end."***
>
> TIM MCELREATH
> Director of Technology
> Discovery Network



"There was a pretty popular weather chatbot and people were sending it images. Why are you sending an image to this bot for weather? They didn't realize until seeing the analytics that showed all the different content types and what that actual content is and so they saw there were a lot of images coming in and so they decided to just to respond to it was kind of a quirky, fun response and what that did is they started to see engagement go up. Come to find out people are sending the images because they treated the bot as if it was their friend and you send images to your friends. So, they also send images to the bot."

ARTE MERRITT
CEO, Dashbot

Use of Images and Symbols

Text isn't the only visual aid that is useful in communication. Images are also an option in some chatbot conversations. You can upload an image to express a thought, provide additional information or move the conversation forward. Sometimes, this is about adding to the conversation and in others it is about connection between the human and the bot.

This isn't limited to actual images. Keyboards often support emojis as inputs as well. In turn some chatbots can send images back to users to create a richer experience. This obviously is not something a voice app can do. So, when considering whether you can port your chatbot design over to a voice app, you have many differences to take into account. With that said, there are many similarities, particularly in the benefits to users of conversational interfaces.

Key Chabot vs. Voice Apps Considerations

It is tempting to treat all conversational interfaces the same. Chatbots and voice apps certainly have many features in common. They do not force users to employ programmatic interfaces to adjust to the capabilities of the system, but instead enable them to communicate by typing and speaking human speech. The system adjusts to the user's natural methods. However, there are many important differences between chat and voice that user experience designers must keep top of mind.

1. Consider how the conversation will change when it must be synchronous
2. Consider if the lack of a text transcript may require a different conversation flow
3. Determine if you need to change the order of your messages so the key point in the response is at the end and not the beginning
4. Understand that voice and audio-only interaction may limit the type of information and connection you can make with a user

How Voice and Visual are Different

The dominant computing user interface is visual. It involves visual navigation and visual response. The web, mobile, smart TVs, automobile dashboards and more all tend to default to visual first and often visual-only interaction. We all understand intuitively that voice and visual interactions are different and while sight and sound can be complementary in many regards they are opposites. It may go without saying that you need to treat these UIs differently. However, it is instructive to discuss their differences particularly because multimodal development requires them to work in concert while also supporting their operation individually.

Visual Requires More Attention

The first difference to highlight is that visual interaction requires more attention of the user. They must concentrate their eyes on the interface to know how to enter inputs and consume the outputs. Often, visual interfaces also require manual input to start or extend the process. This is as true of a video game as it is of a chatbot. Eyes and typically hands are required. Voice can be a hands free and eyes free experience. The user may be giving attention to the voice interaction, but their eyes and hands could be otherwise occupied. In fact, some of the first popular Alexa skills were based on recipes and designed to be accessed while cooking.

So, one of your first considerations when designing for voice should be whether the user must apply focused attention to interact or if they can be multitasking. It is not to say that visual interfaces don't permit multitasking. However, multitasking in visual interfaces generally requires more context switching while a voice and audio user can maintain an interaction while simultaneously using their eyes or hands for something else.

Voice Comes with More Variables

Another difference is that voice comes with more variables to track and more variety to support. Visual interfaces are programmatic and by definition establish boundaries around what a user can do. If the function is not on the screen, it cannot be done. That means you have more consistency on what users are doing in terms of inputs. They can change the input sequence, but cannot go beyond the inputs available. Voice is unbounded in terms of inputs. A user can say anything. The burden on the voice UX designer is to understand widely varying inputs, translate those accurately into intents and then determine how best to fulfill the request or move the conversation forward.

VOICE VS. VISUAL

LISA FALKSON, Amazon

"[The difference of] voice versus visual has to do with the temporal aspect to voice. This has always been the case whether you're on the phone or you're in a conversation. There's a reason that we take notes in a class lecture if we don't record it. It's because our memory for long strings of audio is not the same as it is for writing something down and then you can just look at it over and over again. So, when you're designing for voice, you have to be aware that you can't talk and talk and talk and talk. For example, a legal disclaimer should not be read out loud unless it's 'your call may be recorded.'"

UI Efficiency Differences

At a design concept level, visual and voice introduce different benefits and constraints throughout the user interaction. Visual interfaces are very good with simple inputs and can handle complex or simple outputs. However, they tend to be inefficient with complex inputs such as compound queries and the presence of modifiers. By contrast, voice is very good with both simple and complex inputs and does well with simple outputs. It is complex or lengthy outputs where voice struggles.

> *"Voice tends to be really efficient for input, because it can handle complex tasks that take multiple steps through one-shot queries. It can get you deep into what it is that you're looking for with a single query, instead of visually having to go find things step-by-step," Karen Kaushansky*

These strengths and weaknesses should influence your voice UX design. There are some outputs that will simply be too long or too complex to effectively communicate through an audio response. That means you may need to spend extra time formulating

> *In the web and app world, you have 100 people that do one [specific] thing, and you optimize the flow for it. In the [conversational world], you have 100 people that do 100 different things in infinitely different ways, so everything changes. The analytics change, the way you have to think about design changes, the QA process you go through, and the expectation of the consumer.*

SHANE MAC
CEO, Assist

concise answers or developing a flow to take the user to another channel for the output.

> *"If you have a five-step flow to book a hotel, and I say "hey, I need hotels on Saturday for three nights," that is now two steps less than the web, because language made it shorter," Shane Mac*

Key Visual vs. Voice Considerations

1. Understand that voice users may be multitasking and not paying full attention to the voice app.

2. Voice requires more time considering the variability of inputs that a user could employ to access the voice app features or attempt to access features that are not present. This typically is not an issue with visual design because it places clear constraints on user input.

3. Voice is great for almost any type of input, but it is poor at delivering long or complex outputs. Consider how to shorten responses or switch the user to another channel with visual elements to address complex outputs.

ADVANCED CONCEPTS

TALKING VISUAL CUES
KAREN KAUSHANSKY,
Robot Futures Consulting

"Visuals tend to be efficient for output. On one end, visual displays can provide a lot of space to lay out results and communicate a lot of information at once. Other visual features—for example, Alexa's LED ring—communicate other device states, such as if the device is listening, processing, or if it is online or offline. [Alexa's] LED ring is a key output, even though it feels like this is audio-in, audio-out."

"I think that [visual output] is very important to the experience. Certainly, I don't look over every time that I'm talking to Alexa, but there are key times I look over to make sure. It's reinforcing. It's like if you're having a conversation with a person and they aren't looking your way. This [LED ring] is the response that we've both started a conversation, that we're both here and present, waiting to move the conversation forward."

Multimodal Design

Voice-first doesn't mean voice-only all of the time. Voice interaction has quickly broken out of the plastic cylinders also known as smart speakers to inhabit smart displays, smart TVs and new, more dynamic mobile interfaces. Multimodal design principles suggest you should design experiences that are consistent across devices, continuous when moving from one device to another and complementary between modes of interaction according to Jan König, co-founder and CEO of Jovo:

> *"Every device has its own context. Context-first is delivering the right information at the right time on the right device."*

The idea of voice-first is to ensure the designer starts with the more complex interaction model. Visual interaction is well understood, but can crowd out voice interactions and indeed lead to decisions that make voice-only interactions all but impossible. User experience designer Karen Kaushansky says you should give users control over how they want to interact. If they are in a position where voice is their only option or simply the preferred option, you must begin with an understanding of how you can deliver a voice-only experience, but you shouldn't limit yourself. She characterizes this as customer-first design:

> *"Customer-first, not just voice-first: Giving users the choice of how they want to interact, how and where and what inputs they want to use, is becoming part of the natural landscape of design."*

Visual Offers Different Benefits and It Isn't Just Pictures

When incorporating visual information into a voice interaction, it can significantly enhance the user experience. Some information is simply more quickly and effectively delivered through visual interaction. You can show a user the high and low temperatures for the day or the week in an instant on a screen. It takes several seconds to read off the information for the day and can take a minute to convey the information for the week. Delivering the information through both images and audio provides flexibility for the user.

It is also important to recognize that visual isn't just pictures. It might be an image or video that complements a voice-driven interaction or it could be an LED that provides the user an important signal about the state of the interaction.

Then there are avatars. This is a visual element that doesn't just complement the voice-driven experience, it can transform the interaction to be more humanlike. You don't have a conversation with a video, gif or still image, but you do with a voice and the expression options multiply when you use an avatar.

> *"If your multimodal design includes an avatar...that allows you extra means of communication through body language or eye gaze, for instance that can also enhance the experience of the user," Cathy Pearl*

Choose Multimodal Elements Based on Use Case and Device

Cathy Pearl also says that deciding how or if to use different multimodal elements should be driven by your use case. Not every user scenario demands verbose communication or a protracted back-and-forth discussion.

> *"In deciding what role video and visual displays will play in the expansion of voice assistants, it depends on the use case. If you don't need it: don't overcomplicate. For example, with banking or something similar, just use voice. If you're doing something like story-telling or something that is emotionally heavy, then using an avatar or video could be really important," Cathy Pearl*

Tim McElreath of Discovery Communications points out that your multimodal mix should be influenced by the use case and the devices consumers are interacting with:

> *"Audio-only allows you to be more verbose than if you have a visual component. People can get very impatient if they're processing information both visually and through audio. People process information a lot more quickly through a visual medium than you can tell them that information through audio. You need to architect systems that can take into account device capabilities."*

In other words, you might need to modify your response based on both the use case and the device. If you know the devices offers visual display options, you can then deliver those to a screen. You may also want to deliver different audio content if you know a screen is present. Without a screen, it may be necessary to be verbose to fully communicate the requested information. The presence of a screen and visual elements may mean you can deliver more concise audio knowing that the user is also consuming information visually. Noelle Larchartie, cognitive services lead for developer experience at Microsoft and former machine learning developer lead for Amazon Alexa, says there are also many responsibilities from the NLU perspective that mirror these design choices.

> *"Let's say we come out with a new function like a screen. Now all of a sudden the word 'show' becomes important. I can't just say play something that has video and audio. So, we now have to work on our model to make sure that is accommodated."*

Key Multimodal Considerations

1. Voice-first doesn't mean voice-only. It means you need to design for voice first because it imposes the most constraints when you know some use cases will be voice-only due to device limitations such as on smart speakers. If you start with multimodal and don't adequately account for voice-only scenarios, you will miss the mark.

2. Multimodal requires you to develop complementary elements where voice and visual work together. Don't simply replicate a side-by-side voice-only and visual-only experience to be run simultaneously. This means you may need to consider multiple use case configurations, support each of them and deliver variable experiences depending on the user context. Give users a choice of how they prefer to interact.

3. Multimodal isn't just about images and video. The Amazon Echo light ring and Google Home spinning light dots offer multimodal feedback. Avatars can imitate human facial expressions and body language. And, don't forget text. Even if most use cases may predominantly employ a single mode of interaction, designs will be required to support multimodal when available.

Multi-Context and Multi-Channel Design

Bill Gates famously wrote in 1996 about the future of the internet stating “content is king.” Content is critical when it comes to voice UX but Gates was referring to the internet as a revolutionary digital publishing and distribution technology. With conversational technology, “context is king.” Whether a particular content artifact is great for a user at any particular time depends on the context. What is the user trying to accomplish? What device or devices are they using at the moment? Where are they?

Designers must understand that accurately determining intent and then fulfilling it can differ radically. Identical user requests from the kitchen or the car involve different assumptions, disposition and constraints for the user. Multimodal is more about optimizing the content for the use case and device. Multi-context from a voice UX design perspective involves understanding the user’s situation at any given time. The better designers can match context with the appropriate conversational elements, the more successful the UX outcome.

Accommodating Multi-Channel

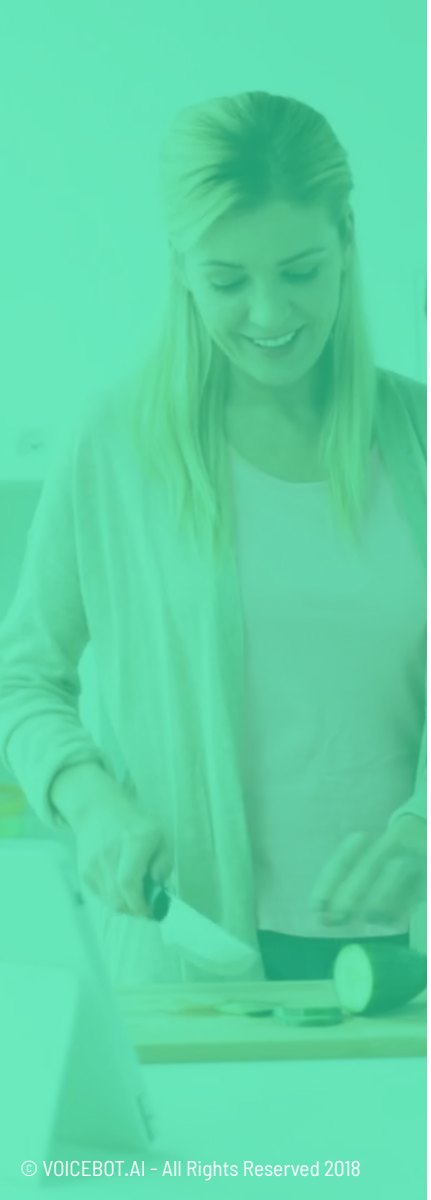
A favorite term in marketing today is omnichannel. The idea is that marketing should be available and be consistent across all consumer channels. Voice is complicated because today it embodies both a new channel, namely smart speakers, and an interface extension for existing channels such as mobile, laptops and IoT. If you

THINK CONTEXT FIRST

JAN KÖNIG, Co-founder, Jovo

“[Multi-device user experience] is all about context: learning where the user currently is, with all the many devices that are available to them today; understanding, if the user is just sitting on the couch, tapping on their tablet. Or, is the user on the go, with a bad internet connection on the subway.”

“Every device has its own context. Context-first is delivering the right information at the right time on the right device.”



want to be available to consumers through the smart speaker channel, you must build a voice app to simply have presence.

However, designers should consider that their voice app may also need to work consistently on smartphones, smart TVs and in the car. Will this be one voice app for all surfaces or several that are optimized for each user scenario? Smart speaker-only initiatives may treat voice interaction design too narrowly.

> *"Google Assistant, and Alexa, and all their ambitions, and all the things they are, and Siri and Apple—they all have the same vision, it feels. And I think that's why you see such a massive, massive hype in the space. They're all going to think the screen should talk to the phone and talk to the chat and Alexa's on this and it's everywhere. They all have the same thing—that it needs to be everywhere," Shane Mac, CEO, Assist*

This is a finer point on the multi-context discussion. Context may be king but content and capability are members of the royal court. The questions for designers are, where else might this experience show up, are we prepared to accommodate variability to support alternative channels and how should these considerations impact our assumptions? Voice will soon be everywhere because it offers new value even in old channels.

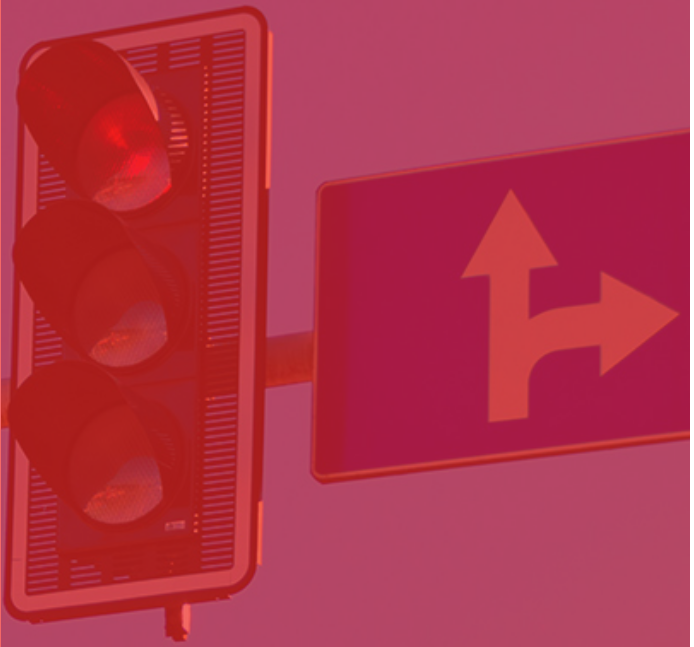
> *"It's not necessarily that [voice] is a content channel itself, but it's a conduit between two different content channels that removes a significant piece of friction for the user," Tim McElreath, Director of Technology, Discovery Network*

> *"Multi-channel design is geared towards brands and enterprises; continuation of conversations from one channel to another; create [applications] that know that someone might be moving from one channel to the other, but making that a continuation and not a stop and start," Karen Kaushansky, Consultant, Robot Futures Consulting*

Key Multi-Context and Multi-Channel Considerations

1. Determine the variety of situational contexts that users may have when using your voice experience and which of those you are prepared to support. Then evaluate what modifications you could build into your intents to deliver a context-dependent experience for users and how you will determine when each context is present.

2. Consider all of the channels that your voice app may be experienced by users. You don't need to build for every channel at the beginning but design with the end in mind and consider how flexible your voice app experience needs to be to support multi-channel use cases.



The Power of Testing and Tuning

Testing is a critical step in the design process for voice apps and not just the build process. It should be executed throughout development and prior to release, but to stop there would be to stop short. The logs of how users are interacting with your voice app supply critical information that is often not apparent prior to release. There can be eye-opening gaps in the software, places where the conversational UI fails, and opportunities for development and improvement.

Shane Mac of Assist says that you must understand that errors for voice apps are what he calls backwards. In traditional apps, errors are to be avoided and indicate something in the implementation went wrong. When you find errors, it typically means you need to close off an option so users don't inadvertently stumble into the error in production. For voice apps, errors show you what was missed and how the design can be extended.

> *"You have to think literally differently about the software that it never has all the answers and learn every day...I think even feedback loops are hard. It's hard to get feedback every day that you're wrong...And that's gonna be a hard shift to go through for a lot of people that give lip service to launch and learn but they're really uncomfortable with it," Shane Mac, CEO, Assist*

Design Doesn't End with Launch. Review Logs for New Ideas.

Some voice app developers have embraced this idea that errors offer insight into how to create a more robust and extensive user experience. Arte Merrit from Dashbot has seen this many times and indeed the company's software is designed to show voice app publishers where conversations are failing, often revealing opportunities for both repair and extension.

> *"The Ansible IPG media guys built Google Actions and [Alexa] skills for Kia and Miller Coors and... they keep looking at what people are saying or asking and adding content. I think with Kia when it started out it was just tell me some brief stuff about Kia like where can I buy it or what's the price. Then they saw based on looking at the utterances what people are saying and they added additional things [so it] almost became like an online voice manual for the car," Arte Merritt, CEO, Dashbot*

Logs Will Also Teach You How to Tune Your Voice App

Cathy Pearl, author of Designing Voice User Interfaces, agrees that logs are critical for many reasons. With voice apps, there is always a certain amount of uncertainty until you have users running into errors while in production. Pearl points out that the secrets to improving your design and optimizing your conversation flow are in the error logs:

> *"[Because logging and tuning is so important,] we had a whole team dedicated to looking at the data we got back from these IVRs. It always surprises me today when someone is building a voice system and they do not realize they need to be looking at their logs, and what a wealth of information [these logs contain]. [This information] is so important, especially for voice, to make your systems good. Anytime you are doing a brand new thing—like symptoms or some other category of items—the logs give us much more intel to boost the match rates with. I would say one of our most dramatic ones was a place where we boosted the match rate by 30% based on our analysis and tuning—it can make a huge difference."*

It can be problematic when you don't have access to the full utterance information according to Pearl. If you can't see the nature of the failures, then you don't have much to go on when trying to make updates. When everyone controlled their own installed voice recognition systems this wasn't an issue. However, with the emergence of consumer-facing voice platforms, the data is not owned by the voice app publisher and utterance detail is not always available.

> ***One of the problems with Amazon Echo is that you can't look at those logs. I think a lot of skills end up suffering because of that. Sure, you can see when someone matched to a particular slot, and that's great, but you can't see the failures. If you don't look at the failures, you really can't improve and perfect your system.***
>
> CATHY PEARL, Author
> Designing Voice User Interfaces

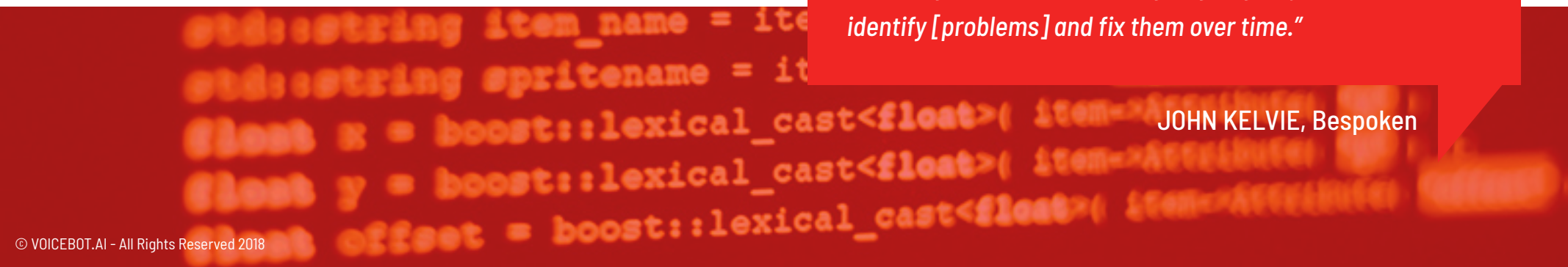
Unit and System Testing Will Reveal Common Errors

As important as ongoing testing and log analysis is in production, the first step is unit and system testing, sometimes called end-to-end testing. This is where common errors will be revealed and cleaning them up early will save you a lot of headaches before you go into live user testing or launch. Bespoken supports over 1,000 voice apps in production and John Kelvie shared what his team has learned about testing in Voicebot Podcast Episode 55:

> *"The most common types of errors are what I would say are really basic interaction model and speech recognition errors. So, that's the first thing that people need to chase out and to make sure that they really are testing for. The second most common thing I would say is if you get into having a more complex interaction model and you get into using the NLU more, you are just going to see very unpredictable behavior. That's where coming up with lots of different test cases [is] really going to help you."*

Kelvie also went on to discuss the added challenge of launching voice apps for multiple countries. This may appear obvious that you need different testing for different language given the variance in vocabulary and grammar. However, Kelvie also pointed out that you still need to do this for variants of the same language. A particular problem that can arise is how the voice assistant natural language understanding (NLU) system interprets idioms.

> ***"I would say for the people that are doing things across different languages and worldwide and dealing with issues around idioms, phrasing for languages themselves...even if they have some sort of QA team, they probably don't speak Japanese. And if they speak Japanese, they probably don't speak Italian. If you're launching into those markets...you're going to have things that are broken and you just need to get a test regimen in place that's going to give you a chance to identify [problems] and fix them over time."***
>
> JOHN KELVIE, Bespoken



Avoid Problems by Using Real-World Testing Before Launch

Unit and system testing are essential. Mistakes happen in the development process and these test practices are the most efficient way to identify common errors. However, these methods also tend to find errors in the scenarios that designers and engineers expect users to experience. The test use cases tend to bias toward validating the design and not revealing unexpected errors.

> *"There are some things clients go into testing thinking about. These are known flows and utterances they are looking to verify, but there's always the 'wow, we didn't even think about that,' issues that weren't thought of until after the testing results come back. As there's not always a clear path to getting to where the user needs to go [in a voice app], there is a huge focus on navigation and UX. Unlike mobile apps and the web, the user has much less of an idea of what's next in the flow," Charlie Ungashick, CMO, Applause*

As a result, testing should be inclusive of people far outside the build team. Bring in people who will actually use the voice app and who do not have the same technical knowledge as a development team or any preconceived notions about the intended user experience. Real-world user testing is the only way to fully test a voice app before launch.

> *"During user testing and design, bring as broad a sampling of people in to help design and test your systems," Tim McElreath, Discovery Network*

> ***Our goal is to replicate the end-user experience as closely as possible during testing. All of our testing is done with real people, voices, and devices – under the same use cases and in the exact environments where they will be used by end-users. Testers could be testing at home or on the move – wherever they would typically be using Alexa. In the case of Alexa skills, our community is testing by speaking to their personally-owned Echo devices and other AVS-enabled devices like vehicles and smart speakers.***

CHARLIE UNGASHICK
CMO, Applause

REAL WORLD LESSON: INTERPRETATION ERRORS

Dylan Zwick, CPO & Abhishek Sutan, CEO of Pulse Labs

"Make sure you spend a lot of time focused on the distinction between you didn't understand what the user said and you're unable to do what the user asked. If you're going to respond for example, 'I'm sorry, I didn't understand that,' or 'I'm sorry, I don't know how to do that,' make sure that you're prepared for situations where the user might say something that you can do just in a slightly different way and you might respond. 'Oh, I can't do that.' Or, the user might ask for something that you can't do but you'll say "Oh, I'm sorry, I didn't understand that."" Dylan Zwick

"That actually tends to be extremely frustrating for users because essentially the skill is not able to handle that input but comes back and tells the user that I don't understand you and the user just keeps saying the same thing, just trying to change the way they're saying it." Abhishek Suthan

Another type of error is telling a user that you cannot answer a question or perform a task, when the voice app actually can. These misinterpretation errors along with misunderstood speech or unsupported requests are the types of errors that are far easier to unearth when you have a group of real-world testers banging on your voice app. So, the best practices include some simple rules. Test throughout the build process. Conduct real world user testing before release. And, continue testing in production so you continue to improve and upgrade your products.

Key Testing Considerations

1. Conduct unit and system testing to identify the most common build and design errors.

2. Conduct real world testing that replicates actual user interactions to reveal misinterpretation errors and situations not considered during design.

3. Update your design before launch to accommodate user interactions that are not at all or are not well supported in the initial design.

4. Review errors in your production logs to determine where the design needs to be modified and where you should consider extending the content or capabilities of the voice app.

Branding in Voice

Most brands have strong visual elements today but are entirely lacking when it comes to an audio or conversational personality. Defining the objective of your voice app may be step one and audience targeting step two, but branding is an immediate step three and ideally precedes all of these activities.

> "Adding things like non-verbal audio—so sound design—the things like sound effects and 'earcons,' which are the *audio-equivalent of icons, jingles, opening music, how do I make this feel like my brand.* All of those things should be defined before you ever write a line of code," Mark Webster, CEO, Sayspring

It certainly should come before you get too far into design because the brand's voice will dictate many of your choices. A good place to start is your brand persona says Pat Higbie of XAPPmedia:

> "What we encourage brands to do is not only get there early and create a presence but also to create a voice persona...a way that people can identify with your brand audibly and a sound-scape along with that. If every brand out there uses Alexa's voice on Amazon and the Google Assistant voices on Google Home, then what is your brand? What are your audio assets? *So, it's super important to create a voice persona for your brand*...to actually make your content come alive and make it believable to your listeners."

Shane Mac zeros in on the element of the persona that makes voice different from chat. The words, phrasing and discourse matter, but the voice itself, the sound, makes it different.

> "Brands are definitely going to have to make their voice. Take Lonely Planet for an example. They are such a strong travel brand. In people's minds, they have a voice. They're from Australia, probably has an Australian accent. It's a female—the way they're talking about it. That layer is what is completely different [between text and voice chat]. *The tone of the words isn't the game. It's the actual voice.* And I think that's where [chat and voice] completely diverge. If the cores are the same, but the output—how it sounds—is different, then that's just something we think as a layer, not a different platform," Shane Mac, CEO, Assist

Focus on the Language

However, that is not to say a brand can only convey distinctive elements of its persona by using a voice actor and recorded audio. Words matter both in selection and number. A voice app that responds with succinct, direct language will convey a different personality than another that uses flowery language and lengthy prose. However, that is not all. Words on a page will be rendered as you see them. Words read by a voice assistant or even a voice actor will vary based on punctuation and other markers.

> ***"Getting down into the details of design, of what is the actual language that's part of the response—in the world of voice interfaces, the way you place commas and periods and spaces will change how these voice assistants say the speech back. So, making sure to be specific about how you style that language becomes really important; being able to go through pronunciations—is our brand name being pronounced the right way?***

MARK WEBSTER
CEO, Sayspring

Consider the Channel. Be Consistent but Not Rigid.

Many people believe the first rule of branding is consistency. The thinking goes that visual and textual language should be consistent or there is a risk that the consumer will misinterpret the brand. This rigid interpretation can cause a significant problem. Marketers that repurpose content intended to be read and port it unedited to a conversational interface quickly divert from the brand's personality. Humans do not speak as they write. This diversity of communication doesn't make that human any less authentic or distinct. They are simply optimizing their language for the mode of communication.

> "Overall, we're seeing so many different channels…we're trying to step back and think, from a product perspective: *what is our brand voice across all of these platforms? How does it change from platform to platform?"* Tim McElreath, Director of Technology, Discovery Network

It is not just that conversational communication is different from marketing copy. There is also the issue of the breadth of actual conversation. Even if you want to repurpose content already available in written form, get ready to create more content. Written content assumes one-way communication. Conversation presupposes two-way communication. Some of the responses will be short while others will require more exposition.

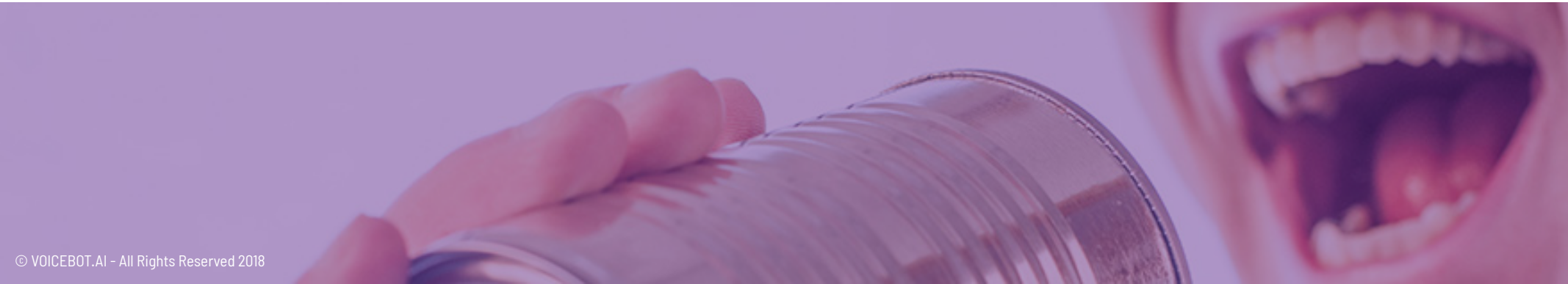
Also, recognize that some information ideally communicated in visual and textual form does not translate well into voice. Mathematical formulas, scientific representations of elements and dense information all immediately create issues with user comprehension and retention when delivered audibly. You may decide that conveying some of this information is best handled by shifting your user to another mode where they can visualize or read the output. In that case, your audible communication may be brief or simply refer the user to another channel. When considering your brand persona and the limitations of the voice medium, you will necessarily diverge from some current elements of your brand. Those efforts will be in service of actually conveying your brand accurately through voice.

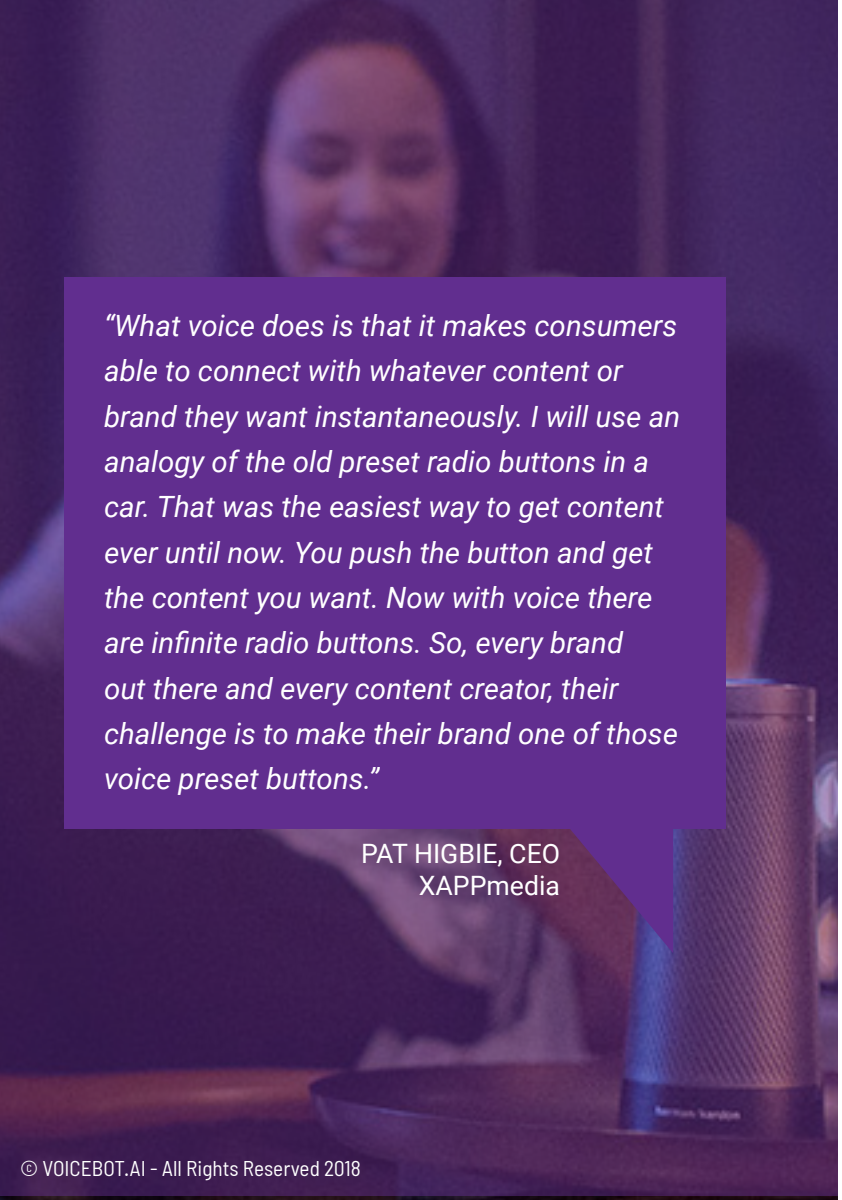
CHANNEL CONSIDERATIONS

CHRIS MESSINA, Product Designer

"You essentially have a set of relationships with entities in the world: some of them are real people, some of them are virtual entities—for example, your bank—some of them are virtual services, like Uber. And it's important that you're able to access those things from multiple different contexts. Each channel needs to know who you are, and needs to be responsive to whatever the most recent context might have been. And you feel the frustration when you try to move from one context to another, and the service, the brand, the person—whatever—isn't available on other contexts."

"I think there is a recognition that brands need to think of themselves as conversational brands that exist in lots of different places. And the more that you're available and accessible at a moment's notice, in the format that your customer expects or uses the most…you just need to be present in a lot of these contexts."





"What voice does is that it makes consumers able to connect with whatever content or brand they want instantaneously. I will use an analogy of the old preset radio buttons in a car. That was the easiest way to get content ever until now. You push the button and get the content you want. Now with voice there are infinite radio buttons. So, every brand out there and every content creator, their challenge is to make their brand one of those voice preset buttons."

PAT HIGBIE, CEO
XAPPmedia

Key Brand Voice Considerations

1. Define your brand persona and how that will translate into an actual speaking voice.

2. Determine what other sounds beyond the speaking voice will be associated with your brand. Music, jingles, earcons, background noise and other sounds can all be used to enhance the user experience and make your brand come across as audibly distinct.

3. Pay attention to the words, phrasing, length, volume and other factors. This is true when coaching a voice actor and doubly important when using the voice assistant as your voice. In the latter case, the words and configurations you make in the code to optimize pronunciation, tempo and other factors will be the primary way your brand essence is conveyed to the user.

4. Write for conversation. You cannot properly express your brand personality by regurgitating the copy from your website. Written text won't translate well into audio communication and will have many gaps and misalignments when it comes to actual conversation.

5. Optimize for the channel. This is a finer point on the idea of writing for conversation. You need to ensure your personality is aligned with the constraints and expectations of the channel. The first websites for many brands were recreations of their printed brochures. Among the earliest mobile apps for brands were attempts to recreate elements of the corporate website. Initial forays into social media saw brand marketers copying text from ad campaigns and official materials. All of these approaches did not account for the unique aspects of the channel. Voice is different and you can only convey your brand properly when aligning content with the medium.

BEST PRACTICES, COMMON MISTAKES & RESOURCES

Best Practices to Embrace and Common Mistakes to Avoid

There are several best practices to embrace in your voice UX efforts and just as many pitfalls to avoid. We have summarized a few of the key concepts to keep in mind when you start your next voice project or are looking to take your game up a level with your current voice app.

5 Mistakes to Avoid

1. Assuming the user is like you

Raluca Budiu, director of research at Nielsen Norman Group, wrote in a 2017 article, "You Are Not the User: The False Consensus Effect," that, "The false-consensus effect refers to people's tendency to assume that others share their beliefs and will behave similarly in a given context...users have different backgrounds, different experiences with user interfaces, different mindsets, different mental models, and different goals. They are not us." We all talk, so it is natural to think that we know how a conversation will flow and this leads to many voice apps that are too rigidly designed. It is one of the key failings of many of the early IVR systems. Tim McElreath of Discovery Communications has overseen the development of some of the most successful voice apps on Amazon Alexa and puts it this way.

> *"Don't expect that people talk the way you talk, or converse the way you converse. It is crucial to assume that there's going to be a diversity in the way people converse even within a single language."*

2. Overlooking the importance of prompts

It seems logical to focus first on the key content and dialogue you intend to offer. However, this often means that the prompts--the questions you ask to help move the conversation forward--are neglected. Cathy Pearl from Google says that is a mistake.

> *"A lot of people forget that the prompts themselves—that is, what the voice system says and asks the user—is a crucial, crucial part [of voice design]. You need to spend a lot of time crafting those prompts. A lot of people think of it as 'the icing on the cake,' but there's a skillset involved in choosing exactly how to craft those prompts."*

3. Relying on words alone

Voice assistant interactions offer the opportunity to introduce sounds beyond language to enhance the user experience, increase comprehension and enrich the interaction says Sayspring's Mark Webster.

> *"One [mistake I see often] is not enough use of non-verbal audio. When I open your 3rd-party skill, do I hear a tone or a jingle or something that lets me know that I'm now in this experience. Or, is there another interaction with your brand where that pulls together—something like the NBC chimes or the Intel inside. Is there some signature audio that is part of it? When something works, do I hear a success 'ding.' When something doesn't work, do I hear a sort of buzzer sound letting me know it doesn't work? That's a big one."*

4. Assuming text in logs conveys clear meaning

People assume you can understand user intent simply by reviewing text transcriptions of utterances in the logs. That is incorrect according to Tobias Goebel of Sparkcentral. This data is helpful, but there are variables of expression and context that can alter the meaning of text.

> *"Language can be illogical when the words you utter do not reflect, at all, the meaning. The obvious example is irony or sarcasm, where you don't have enough signal in the written word to derive the true intent [of the statement]. So a mistake that many [people] make today is assuming that just by looking at the raw text from a message, that you can interpret, one-to-one, the meaning of that message. The reality is that you cannot. [This applies] not just for reasons of irony which is an extreme example, but also for reasons of context—the contextual situation of where somebody utters a sentence of types a message influences what the meaning is."*

5. Assuming unit and system testing are sufficient

Another common mistake is that voice app developers do not conduct sufficient real-world testing. Charlie Ungashick from Applause points out consumer experience risks that may not be apparent in design or surface during traditional system testing.

> *"Consumers expect to be able to launch the skill, speak naturally, and accomplish whatever they set out to do. The problem is that conversations are highly variable, and there are thousands of ways to say the same thing – especially when you consider different languages and dialects. A leading auto manufacturer works with us for Alexa Voice Services testing. Our testers discovered that Alexa really struggled to understand US drivers with specific southern and midwestern accents – which for this client, represented an important demographic…Crowd testing, where we're using real end-users from across the globe, is a natural fit to uncover and account for this variability between people, geographies, and devices."*

5 Voice UX Best Practices

1. Set expectations and don't over promise

The commercials and hype surrounding Siri's launch on the iPhone 4S created an appearance that the new voice assistant could do anything. It was often portrayed as similar to having a human assistant and Apple's reputation for innovation made the claims seem plausible. But, they weren't. The first release of Siri was amazing in many ways but its performance could not match the hype. As a result, many consumers harbor negative associations with Siri's capabilities to this day despite its remarkable improvements.

By contrast, Amazon launched Alexa saying it could do a few things, but would learn over time and be able to do more. Expectations were set appropriately and consumers fell in love despite limitations of the initial product. Karen Kushansky of Robot Futures suggests voice app developers should set expectations and avoid overpromising. If you say you can do everything, you are sure to fail.

> *"We should design our robots and assistants to be human enough but not too human lest they overpromise. And of course we have to define what "human enough" is…we don't want to trick people into thinking that our robots are human because we're just going to let people down. So setting expectations of what the robot can do becomes huge."*

2. Create a voice persona

Voice assistants can create a challenge in making your brand distinctive. You don't have the colors, logo marks and rich images to immediately trigger an association or emotional reaction from your user. In fact, if you are using the voice assistant provided voice, your audio brand may be undifferentiated from the platform and your competitors. A great starting point is to define your voice persona according to XAPPmedia's Pat Higbie. That can include using a voice actor or simply adopting language, phrasing, timber, pace and style that you will implement through a voice assistant provided voice. It may also include other sounds to make your experience audibly differentiated.

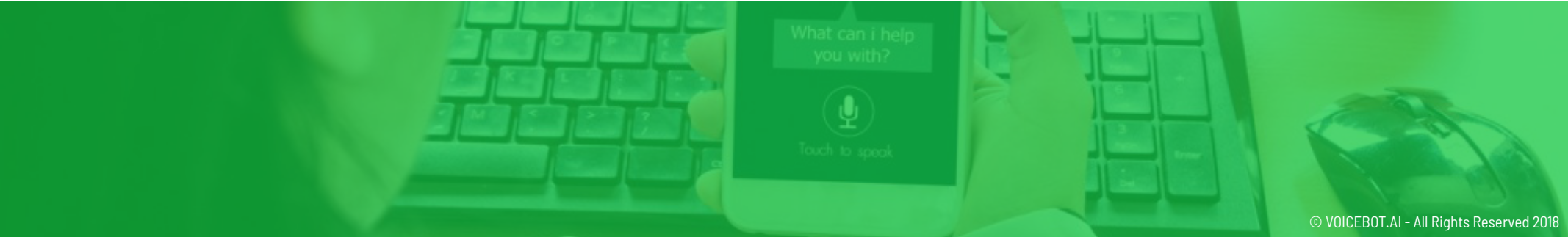
> *"What we encourage brands to do is not only get there early and create a presence but also to create a voice persona...a way that people can identify with your brand audibly and a soundscape along with that. If every brand out there uses Alexa's voice on Amazon and the Google Assistant voices on Google Home, then what is your brand? What are your audio assets? So, it's super important to create a voice persona for your brand...to actually make your content come alive and make it believable to your listeners."*

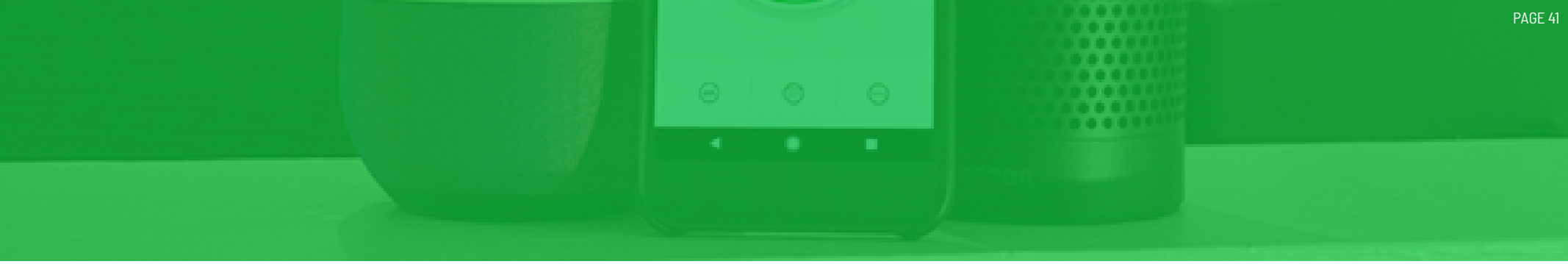
3. Inspect your logs and adjust

You cannot anticipate every error that will arise in a user session. However, you can review user sessions to evaluate errors and determine how to improve your voice app. Include this practice into your voice app launch and sustainment plans. Shane Mac of Assist says that, "everyone in software has been scared of the errors," but the errors help you zero in on where you can correct issues and where new opportunities for engagement exist. Cathy Pearl adds to this by saying:

> *"It always surprises me today when someone is building a voice system and they don't realize they need to be looking at their logs, and what a wealth of information that is. It is so important, especially for voice, to make your systems good."*

Your process should include dedicated time to review logs of real user sessions immediately after launch to understand how you can address critical errors. And, you should plan to review logs on a regular basis so you can fix problems and identify new opportunities.





4. What's important must come at the end

If you are presenting information through text, the most important information should come at the beginning. That ensures the most important point is not missed in case the reader skims the material afterward. By contrast, if you are presenting information audibly the key point must come at the end according to Tim McElreath of Discovery Communication:

> *"If you're reading something off a screen, you want to put the major point [at the beginning]. The big takeaway should come first, because people will read the first sentence and skim the rest. In voice, it's the exact opposite. Whatever you want people to take away from what you've just said, you want to put it right at the end."*

If you don't put the main point at the end, people may forget or even lose the context of the exchange. This is particularly important when asking a question where the response is critical to moving the conversation forward. If you write a great prompt that is designed to gather specific information from a user, but then add other information, the user may forget what the prompt was requesting and undermine your objective. Or, the users may start to answer the question before the voice app has finished speaking and not capture the utterance. Cathy Pearl, author of Designing Voice User Interfaces emphasized this point in her Voicebot Podcast appearance.

> *"What we learned at Nuance in building these conversations was how to structure a prompt so the user knows what they can say when it's their turn to talk and how to not confuse the user. Never in voice [should] the question be followed by the statement, like, 'Which account would you like? You can say Savings or Checking.' Because after you say, 'Which account would you like?' they will say, 'Checking,' before you've said everything and they are going to barge in on the prompt and the prompt is still playing and they are going to get confused."*

5. Design context-first

Don't fall into the trap of thinking voice-only, because so many of our voice assistant interactions are already incorporating visual elements such as images and text. However, you do need to think voice-first because some experiences and maybe most of them are going to be voice-only in practice. So designing for voice first means ensuring the most unforgiving use cases are addressed. This boils down to context-first design according to Jovo's Jan König:

> *"Every device has its own context. Context-first is delivering the right information at the right time on the right device."*

Experienced voice UX designers match the objective and use case with the device context (i.e. its capabilities and limitations) and situational context (i.e. who is using the system, what is their frame of mind and intent, what are their expectations, where are they and other factors). First, ensure you can accommodate an effective voice-only interaction using contextual parameters. Then decide how you can deliver variable voice enabled experiences optimized for different modes of engagement across the device types you plan to support.

Apply Voice Where it Fits

A final point of consideration. Voice is not always the best solution to fulfill an objective. Tobias Goebel of Sparkcentral commented, "The biggest challenge of voice is that it is a synchronous real-time channel... You can't keep [a session] open forever and support a truly natural kind of conversation...You can't really do it continuously in an office. You can't really do it in public, at least in many cases. So, it has its place."

Designers should consider first whether voice is appropriate to deliver the desired impact. If the answer is affirmative that voice should be either the primary user interface or a complementary input, then we hope the insights presented here by some of the industry's most experienced practitioners will help you deliver a first-rate voice user experience.

THE VOICE UX EXPERTS

AHMED BOUZID / WITLINGO

LISTEN HERE
Episode 32

LISA FALKSON / AMAZON

LISTEN HERE
Episode 3

TOBIAS GOEBEL / SPARK CENTRAL

LISTEN HERE
Episode 29
Episode 47

PAT HIGBIE / XAPPMEDIA

LISTEN HERE
Episode 11
Episode 47

KAREN KAUSHANSKY
ROBOT FUTURES CONSULTING

LISTEN HERE
Episode 40

JOHN KELVIE / BESPOKEN

LISTEN HERE
Episode 6
Episode 55

JAN KÖNIG / JOVO

LISTEN HERE
Episode 13
Episode 56

NOELLE LACHARITE / MICROSOFT

LISTEN HERE
Episode 45

SHANE MAC / ASSIST

LISTEN HERE
Episode 18
Episode 54

TIM MCELREATH / DISCOVERY

LISTEN HERE
Episode 28

ARTE MERRITT / DASHBOT

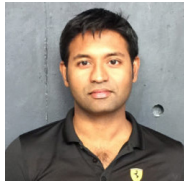
LISTEN HERE
Episode 25
Episode 26

CHRIS MESSINA

LISTEN HERE
Episode 38
Episode 47

CATHY PEARL / GOOGLE

LISTEN HERE
Episode 30

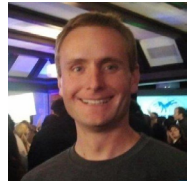
ABHISHEK SUTHAN / PULSE LABS

LISTEN HERE
Episode 40

CHARLIE UNGASHICK / APPLAUSE

READ MORE
Article

MARK WEBSTER / SAYSPRING

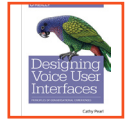
LISTEN HERE
Episode 33

DYLAN ZWICK / PULSE LABS

LISTEN HERE
Episode 40

ADDITIONAL RESOURCES

THE VOICEBOT PODCAST

LISTEN HERE



DESIGNING VOICE USER INTERFACES

LEARN MORE

DON'T MAKE ME TAP

LEARN MORE

VOICE SHOPPING CONSUMER ADOPTION REPORT

DOWNLOAD NOW

SMART SPEAKER CONSUMER ADOPTION REPORT

DOWNLOAD NOW

VOICE INSIDER

SUBSCRIBE NOW

Report Authors

Bret Kinsella
Managing Editor
bret@voicebot.ai

Ava Mutchler
Associate Editor
ava@voicebot.ai

Caroline Kinsella
Contributing Editor
caroline@voicebot.ai

To request custom voice or AI industry research contact: info@voicebot.ai