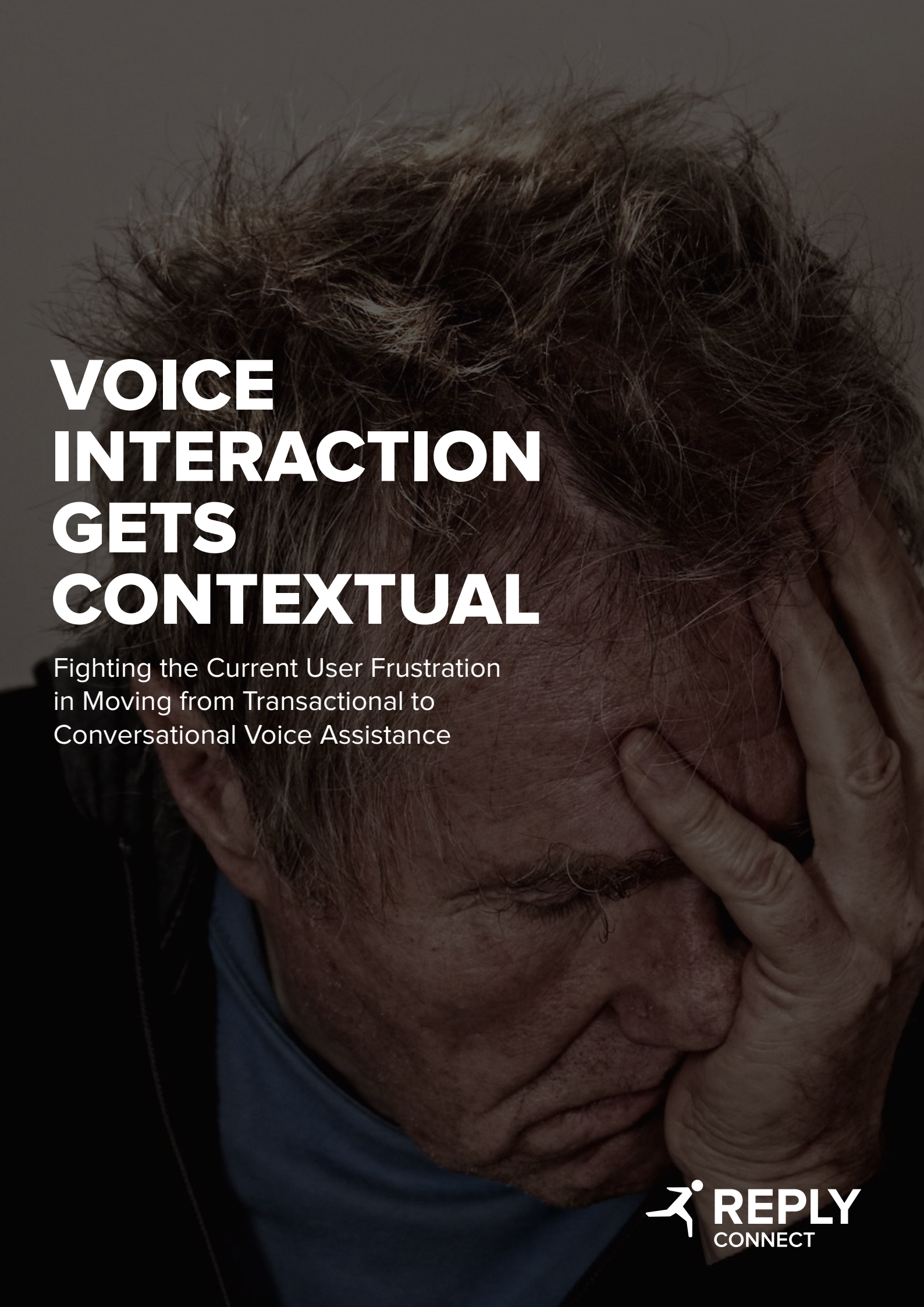


VOICE INTERACTION GETS CONTEXTUAL

Fighting the Current User Frustration in Moving from Transactional to Conversational Voice Assistance

REPLY CONNECT

Why voice interaction?

Daniele Vitali & Gianluca Lupica
Connect Reply's Partners

In Connect Reply we join the physical and digital world, crafting amazing iot technology. **We are experiencing the true beginning of the Internet of Things era, in which the physical and digital world talk to each other and empower people and businesses**. We believe in the Internet of Things as a mean to create a better and more efficient world for people.

We want to free the interactions between people and with the physical world from the constraint of a glass screen, enabling ourselves to become part of the interface with the digital world. Our body and the surrounding environment itself is becoming a profoundly integrated part of the digital experience, and voice user interfaces (VUI) are a predominant part of it.

As we develop new connected products we are working at the real edge of the current technical capabilities of VUIs. Most people look at these systems today mainly in the form of basic automation or very basic passive conversations. Current over the top player investments on AI and natural language processing technologies will in less than two years demonstrate that actual open topic human – machine conversations are doable. Two breakthrough innovations in this space will quickly change our perception: the advent of proactive conversations (where the AI engages first in the talking) and true intent recognition advancements. These innovations will open the world of VUIs to use cases comparable to what today can be engaged through a conversation by a human, and will start driving real services/product conversion and truly innovative commercial strategies.

This is an obvious reality. Yet the world is lacking cases of large brands defining a real voice customer interaction strategy. Back in 2005 very few visionary enterprises embraced cloud, and today are heavily benefitting from that vision. During 2020 we will see the first large brands defining a real strategies, not driven by the fear of missing out but by a genuine will to explore and be one of those visionaries that people will look back in time saying "that was an obvious move".

As Connect Reply, we are ready to take on that challenge with our customers.

Index

Abstract

Like many other design disciplines, the Conversation Designer's role is to identify and validate the problem worth solving by aligning voice with real value and desirability. The goal of the Conversation Designer is to help the user achieve their intended goal. They map out what users can do, while considering both the user's needs and technological constraints. Conversation Designers curate the conversation between user and computer system by writing sample dialogue and defining the flow of the conversation. The Conversation Designer aims to design voice experiences that keep the user's context in mind, and that allow the user to speak naturally. It is important to note that, as voice technology is not quite able yet to interact like a real human, the Conversation Designer can ensure a better experience for users by anticipating every way a conversation may go and reflecting that in their designs. Like in every other design discipline, the conversation designer needs to tell a story, adding a level of surprise and delight to the user's experience.

Index

Introduction	4
Voice: The New Frontier?	5
The State of the Art	8
VUI Guidelines	10
Case Study	13
Tools to design VUIs	18
Methodology & Methods	20
References	21

Introduction

Voice experiences are radically changing the way we interact with technology. With new smart speaker sales reflecting a growth rate of 82.4%, voice is set to be the next tectonic shift in technology. Data from Canalys forecasts that the worldwide smart speaker installed base will exceed 200 million in 2019 and climb to over 500 million by 2023. As voice assistants become ubiquitous in our daily lives, attention must be made towards the usability challenges and barriers that currently exist in interacting with these devices. Voice User Interface Designers must consider the complexities and new challenges when constructing user interfaces, navigation, and informational hierarchies for voice interactions. In order to prepare for the era of voice-first interaction, you must ask yourself what your users expect out of conversations, and what defines a successful user experience in the field of voice interaction.

In order to facilitate the achievement of these aims, this paper will highlight the following research objectives:

- **analyzing the current and future state of VUIs** (Voice User Interfaces), making note of current limitations and future capabilities
- **identifying existing best practices** for designing voice interactions, and defining metrics for evaluating voice experiences
- **drafting a design methodology** for designing voice interactions

By uncovering and leveraging the potential of voice interactions, you can design intuitive, human-centered experiences that can not only delight your customers, but address fundamental challenges they face in their daily lives. No longer will amateur chefs be forced to scrub up, or worse, smudge the screens of their very expensive smartphones, in order to set a timer or to find the next step in a recipe. Using voice controls could be the potential differentiator between an elderly person maintaining their independence, living alone, and needing to be placed in a nursing home.

People want to talk to VUIs just as they talk to humans, and therefore, we must leverage human conversation in the design of Voice User Interfaces. Not only will this allow users to speak naturally to the VUI, it will also mean that users will not have to learn how to use the interface. This paper will explore the importance understanding context will increasingly play in driving proactive conversations, informed by contextual information, which will enrich the experience end users have with a Voice User Interface.

Voice User Interface (VUI)

Voice User Interfaces enable users to access a computing device and complete a task all through speech, instead of using the traditional means of a keyboard, mouse and graphical user interface. Voice User Interfaces take an utterance, or a request that people articulate in a spoken statement, and use intent recognition to understand the action required to fulfill the user's spoken request.

Voice: The New Frontier?

The biggest tech disruption since the smartphone and the very real benefits voice experiences provide—vastly improved accessibility

Voice is considered the next natural step in User Experience Design due to its human form of interaction. Voice is increasingly how many companies, nonprofits and governments are assessing people's needs and delivering services to meet them. After all, we are wired for speech; it is the interface we learned first and know best.

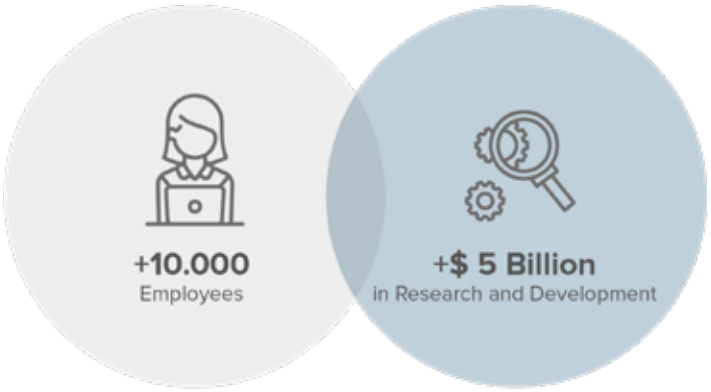
In the case of Graphical User Interfaces (GUI), we often assume that everyone knows how to interact with them. However, users once struggled to learn how to use the Graphical User Interface. It is predicted that in the future, voice will not require any training for the user. Instead, users will simply "speak" as they do naturally. Where users have been accustomed to adapting their own behavior in order to interact with the graphical user interface, the voice user interface needs to adapt itself to the user's behavior.

> "Voice can offer an easier way to search, shop, and multitask. Secondly, it has a lot to do with changes in our behavioral patterns, caused by technological advances, in which we have developed expectations for speed and efficiency, something which voice can offer."

The second era of VUIs

While voice interfaces have exploded in popularity in recent years, they are not new. The first era of VUIs, or computers interacting with humans through the use of voice, began in the 1950s with companies such as Bell Labs and IBM. By the early 2000s, Interactive Voice Response (IVR) systems were thriving with their capability to understand human speech over the phone and carry out tasks. Anyone with a phone could get stock quotes, book flight tickets, transfer money between accounts, order prescription refills, find local movie times, and hear traffic information, using nothing more than a regular landline phone and the human voice. While IVRs could accept simple voice commands, they failed to live up to the promise of a conversational interface, which has hindered their success with users.

We are now in what could be termed the second era of VUIs, where voice assistants like Alexa, Siri, and the Google Assistant have made their way into cars, homes, and into the palms of people's hands — on smartphones. So why the sudden growth in this technology? Well, for one, as Gary Vaynerchuk, CEO of VaynerMedia, puts it: "convenience is king." Voice can offer an easier way to search, shop, and multitask. Secondly, it has a lot to do with changes in our behavioral patterns, caused by technological advances, in which we have developed expectations for speed and efficiency, something which voice can offer. People want a lot of benefit for a little bit of effort. And finally, we expect technology to learn and evolve. With recent advancements in the field



of cloud data and Artificial Intelligence, Google says we are moving from a mobile-first to an AI-first world where we expect technology to be naturally conversational, thoughtfully contextual, and evolutionarily competent.

Investments into voice

Gene Munster of the investment firm Loup Ventures estimates that tech giants, like Amazon, Apple, and Google, are spending a combined 10% of their annual research-and-development budgets, more than $5 billion in total, on voice recognition. Moreover, Amazon's vice president of Alexa, Steve Rabuchin, confirmed that Amazon has 10,000 employees working on Alexa's development. It is clear that these tech giants believe in the assistant as a driver of future growth. In fact, both Google and Amazon have launched investment funds for early-stage startups that are moving the voice and assistance ecosystem forward. Google's new venture fund for investments in voice currently has 10 companies in their portfolio, and there is no cap on the amount the team can invest overall. Amazon's Alexa Fund provides up to $200 million in venture capital funding to fuel voice technology innovation, and has invested in more than 50 different companies. But, it is not just tech giants who are taking note of the future potential of voice technology. There is a reported, "91% of brands investing in voice," according to Adobe. While brands should start investing in voice-first experiences, they must remember to be useful to their customers by taking an intentional strategic approach to the creation of voice experiences.

The power of voice

Speech is one of the first skills we acquire in childhood — and one of the last we lose in our sunset years, long after our vision and motor skills begin to fade. Voice possess four key qualities that make it a powerful tool to use in human-computer interaction. They are:

- **Simplicity**: Voice is about language and conversation, which users have been doing since they were born.
- **Intuitiveness**: Everyone knows how to talk. Even users who are less familiar with technology can reply naturally.
- **Hands-Free**: In some cases, such as driving or cooking, speaking, rather than typing or tapping, is much more practical (and safer).
- **Speed**: As Daniel Padgett, Conversation Design Lead at Google, explains, voice interactions can be extremely fast. Done well, voice interactions can be faster than pulling out a mobile phone from a back pocket. In a world where we are always trying to do things faster, voice is the ultimate shortcut.

With great power comes great expectation. People have inherent expectations when interacting with a VUI, which they developed from oral communication with other human beings. Just as when speaking to another human being, users expect to be understood by the VUI, expect the

VUI to have the ability to carry unbound, back-and-forth exchanges, and have expectations about how the conversation should flow.

Improved accessibility

Based on the four qualities we have listed above, we can see how voice assistants have the potential to provide a type of ease in people's lives for those currently excluded from accessing technology, such as the elderly, the visually impaired, disabled people with arms or hands deficiency, and illiterate people.

As in any design discipline, you must consider your ethical duty as a designer. You should strive for accessibility: designing experiences that anyone can use at any time. Therefore, when designing voice UIs, you should consider limitations, barriers and specific circumstances – such as accents, speech impediments and bilingual households – and design for these individual needs.

While voice is a natural way to communicate, designers and creators of technology must be aware that they often take their ability to be understood for granted. While voice does provide improved accessibility for many people, there are still those who struggle to be understood; millions of people live with speech impairments caused by neurologic conditions such as stroke, ALS, multiple sclerosis, traumatic brain injuries and Parkinson's.

As we move towards an era of voice-first interaction, how can we ensure that those with speech impediments are included? To help grapple with this, Google is using AI to improve computers' abilities to understand diverse speech patterns, such as impaired speech. They've partnered with the non-profit organizations ALS Therapy Development Institute (ALS TDI) and ALS Residence Initiative (ALSRI) to record voices of people who have ALS, a neuro-degenerative condition that can result in the inability to speak and move. Google collaborated closely with these groups to learn about the communication needs of people with ALS, and worked toward optimizing AI based algorithms so that mobile phones and computers can more reliably transcribe words spoken by people with these kinds of speech difficulties.

In addition to improving speech recognition, Google is also training personalized AI algorithms to detect sounds or gestures, and then take actions such as generating spoken commands to Google Home or sending text messages. This may be particularly helpful to people who are severely disabled and cannot speak.

Currently the research aims to accommodate individuals who speak English and have impairments typically associated with ALS, but Google believes that their research can be applied to larger groups of people and to different speech impairments.

***Smart Home of the future. Youtube*, uploaded by Telenor Group, 17 Apr 2018, https://www.youtube.com/watch?v=R2r-QZAcwv8**

The State of the Art

Where voice is now, where it's headed and context as the key to successful conversations

Current State of VUI

SIMPLE COMMAND INTERFACES

Thanks to natural language processing, speech recognition software and A.I., Voice User Interfaces are able to respond to voice commands. However, research illustrates that the technology is not quite there yet for us to be able to use voice to interact with a system in the same way we would another human being–that is, to have a true conversation. Interactions with the VUI system often do not go beyond one turn and revolve around brief question/answer or user-instruction/system-confirmation dialogues.

USER FRUSTRATION

VUI systems struggle with discoverability and conversation repair. Unlike Graphic User Interfaces, which provide users with visuals cues as to what the system can and cannot do, Voice User Interfaces wrestle with relaying to the user what the system is capable of. User frustration develops out of the confusion of not knowing what the system can actually accomplish. Moreover, when the system encounters an error, it often responds with, "Hmmm, I don't know that one." If handled well, designers can rebuild user confidence and trust through repair cases. Designers need to account for things like requests that the skill was not built for and users saying the wrong thing. Errors can present a key opportunity to be a champion for the user, demonstrating that the experience was well thought through.

> "Understanding Context will be the critical component that drives successful conversations."

Future State of VU

FROM TRANSACTIONAL TO CONVERSATIONAL

We predict that within three to five years, advances in AI will make the conversational capabilities of computers vastly more sophisticated, paving the way for natural, flexible back-and-forth interactions. Brands will be able to form human-like relationships with users, and inject interactions with personality and humour. Understanding Context will be the critical component that drives successful conversations.

Contextualized Interactions

REMEMBERING HISTORY

Remembering previous interactions with the user will be the driving force behind Voice User Interfaces anticipating conversation beyond one turn. In order to be more useful and relevant to users, voice assistants need to be able to remember the past: both in terms of from previous conversations and also what was said earlier within the same conversation. By remembering history, the system can streamline interactions based on information shared earlier on, as well as act proactively.

Currently, Alexa can draw on your existing

order history on Amazon— so if you've bought a specific type of AA battery from Amazon in the past, Alexa will allow you to easily reorder that product through a "reorder _____" command. Alexa will ask you to confirm the order, and if you say yes, you're set and ready to go. This illustrates the potential of VUI systems to access and remember information from users' interactions with a service on its other channels.

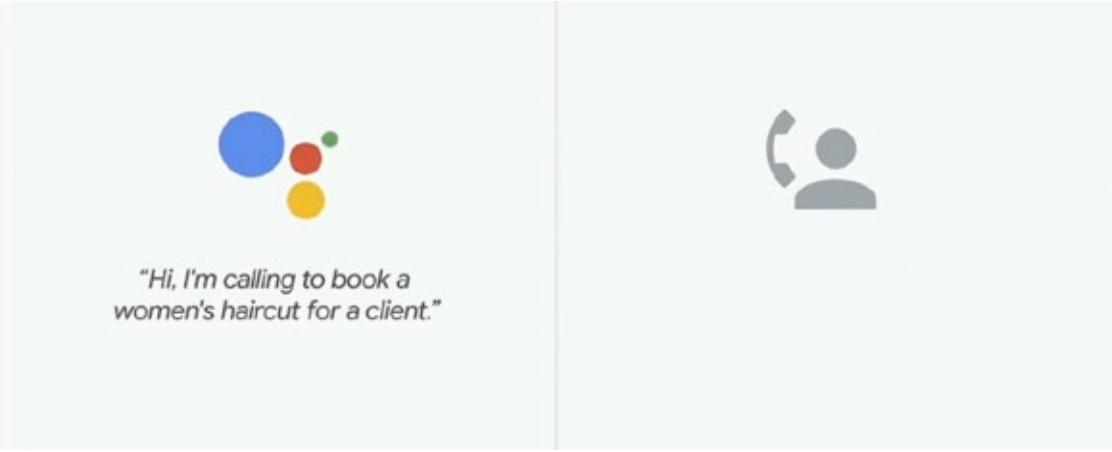
CONTEXTUAL AWARENESS

Contextual Awareness will go one step further to understanding the user and their environment. The system will understand the context of where the user has just been or where they are now, as well the physical and emotional state of the user. In turn, this will allow the system to anticipate user's needs, to sense and respond to users with emotion, and to adapt its behavior and interaction with the user accordingly. Contextual awareness will allow the VUI to adapt its tone of voice in relation to the user's environment. For example, if a user uses the same Alexa Skill at both home, or at the office, Alexa may speak professionally or formally when the user accesses the Skill at the office vs. humorous when the user accesses the Skill at home. Contextual awareness will allow the VUI to understand the user's physical proximity to IoT devices. Alexa has made advancements in context aware Voice control, where if you say, "Alexa, turn on the lights," Alexa knows what lights to turn on based on the Amazon Echo device you are asking. This illustrates the potential of VUI systems to use context aware Voice control in order to allow users to speak to the system naturally.

"The system will understand the context of where the user has just been or where they are now, as well the physical and emotional state of the user."

CONTEXT IN PRACTICE: THE FUTURE IS WITHIN REACH

As we have explored, the future of human-computer interaction is about humans and computer systems communicating in a more natural way. But how do we achieve this level of interaction? At its Google I / O 2018 conference, Google demoed Duplex, an AI-driven voice, which helps people make appointments to businesses over the phone without any interaction from the user. In the demo, you can hear how the AI voice can not only understand the voice of the human on the other end of the call, but it could also respond back with correct answers to that real person's inquiries and questions. Google Duplex's voice sounds quite natural: it responds spontaneously, even adds in filler words like "um", and uses pause breaks, to make it sound even more like a real human. Duplex demonstrates how a future where computer systems partake in the various aspects, nuances and characteristics of human speech, from tonal inflection to mood, isn't far away. To get there, the system will need to improve its intent recognition, its ability to interact, to coordinate, or understand when to insert pauses, and to improve its understanding of context.



Google Duplex demo - Hair Salon. Vimeo, ulpoaded by LangNet, 8 May 2018, https://vimeo.com/268705764

VUI Guidelines

A look at existing best practices for designing voice experiences and how to design for value

Designing for Value

There is only a 6% chance that a person will return to a voice application within seven days of install, according to VoiceLabs. Therefore, as we have previously mentioned, it is absolutely crucial to take an intentional strategic approach to the creation of voice experiences, bringing forth value to users. If the voice experience offers no value to the end user, then there is no sense in creating it. Conversation designers must assist clients in identifying and validating problems worth solving with voice by aligning voice experiences with strong contextual awareness.

RESEARCH PROVIDES THE STRUCTURE

Value creation starts with research aimed at understanding context: who are the users, what are their needs, and where and how they are currently completing tasks and solving their problems. By conducting contextual research, we can uncover compelling use cases where voice interaction can offer a relevant solution.

By understanding how a voice interface can fail, VUI designers can find ways to turn those failures into eventual successes. User trust is built through how the system responds when things go wrong. Brands can increase their competitiveness, and set themselves up for success, by focusing heavily on handling errors and accounting for the variety of utterances users have to try and access the capabilities of the skill, right from the beginning of writing dialogue.

Be Relatable: The Cooperative Principle

Conversations are a cooperative effort. People aim to understand, and be understood by, other people involved in the conversation. They want to give and receive information and they want to influence, and be influenced by, one another. Users hold these same values when they interact with the persona of a conversational user interface, and they expect the persona to have these values too.

Paul Grice found that in order to cooperate in conversations, there are 4 rules, or Maxims, that we passively follow. By being aware of these rules, you can better understand what users expect out of your conversation.

- **Maxim of Quantity:** Aim to make your assistant's contribution to the conversation as informative as is required (and no more) in order to move the conversation forward - not too much, not too little, but just right.

- **Maxim of Quality:** Aim to make the assistant's contribution to the conversation truthful. You should not give information that is false or that is not supported by evidence.

- **Maxim of Relation:** Aim to make the assistant's contribution to the conversation optimally relevant to your users.

- **Maxim of Manner:** Aim to make the assistant communicate clearly, without obscurity or ambiguity.

Turn Taking: The Transmission of Human Conversation

In William Golding's Lord of the Flies, whoever is holding the conch, or a large, milky-white shell, holds the right to speak. This is an explicit form of turn taking and power diffused among people. While you don't need to be this explicit with turn taking in Voice User Interface Design, it is important for you to ensure that the Assistant indicates to the user when it's their turn to speak. Therefore, write prompts that make it clear to the user that it's their time to speak; ending with a question is an obvious indicator, for example.

Additionally, the system should allow users to barge-in after it has conveyed the information necessary for the user to know how to complete the task. Allow users to issue their commands to the system without forcing them to wait for the system to finish talking.

Context is King

We cannot stress enough how important understanding context is in the creation of voice experiences. Conversation Designers must be aware of the environment the user is in and the circumstances they are facing while accessing the service. With this understanding, they can make informed decisions as to which contexts will be most suited to users interacting with voice.

"Most users don't want to talk about private matters in a public space, nor do they want their systems to talk about personal topics to them in the case that others might hear. Therefore, it is critical for designers to conduct research to uncover when and how users are interacting with the service, in order to deliver voice experiences in the right moment."

Conversation Designers must actively focus on the user's physical context and the kinds of conversations they are open to in that context. Most users don't want to talk about private matters in a public space, nor do they want their systems to talk about personal topics to them in the case that others might hear. Therefore, it is critical for designers to conduct research to uncover when and how users are interacting with the service, in order to deliver voice experiences in the right moment.

- Some contexts to look out for during research:
- Hands-busy
- Eyes-busy
- Multi-tasking
- Private Spaces
- "Familiar" shared spaces

Bake Empathy into Voice Interactions

Gary Vaynerchuk, CEO of VaynerMedia, says that "empathy and respect for the end user" is the most important part to creating success in the era of voice-first interaction. Not only will using empathy establish a personable relationship with the user, it will also mitigate user frustration or anger towards using the device; they will even be forgiving when errors arise. The VUI will need to take ownership of their inability to complete certain tasks. After doing so, it can then point the user to corrective actions.

The Graphical User Interface era has shown that no matter how good an interface is things will go wrong. While people are instinctively trying to be cooperative, they are human and they make mistakes. Therefore, designers need to build Voice User Interfaces to be forgiving by designing for conversational repair. Graceful handling of errors is one of the biggest challenges designers face when designing for voice, and it's important to think about it right from the beginning of writing dialog. When things go south, the system should provide valuable options and always aim to get the conversation back on track. Designers should work to anticipate every way the conversation may go, leaving out critical paths will lead to the system responding with, 'I don't understand'.

Natural and Spontaneous

Users should be able to speak to your skill naturally and spontaneously. Make the VUI natural to invoke and natural to remember. Users should not have to think about what to say to the VUI in order to accomplish a task, and allow them to not remember how to say it. They should be able to communicate with your VUI just as they would another human being. Therefore, your VUI should allow people to express themselves in a wide variety of ways with the system understanding a wide variety of utterances. When conducting research, pay attention to the user's language, and the exact words and phrases the user uses to talk about the task that you are designing a voice experience for. To make sure your VUI is able to respond to a variety of utterances, match a variety of utterances to your intents; provide a wide range of sentences, phrases, and words users are likely to say.

> "It is important to remember that there are different types of feedback, in addition to audio, including visual and tactile."

Google Quit Gabbing: Talk Only When Needed

Part of designing conversations involves knowing when to be seen, not heard. Many users of the Google Home Assistant have been turning to internet forums to plead for an option to turn off what they deem unnecessary confirmations. For example, if you ask the Google Home Assistant to "Turn on the lights", it responds with, "OK, turning on 2 lights." Amber Case, Researcher, Author and Design Leader, argues that we should be aiming to create calm technology, or designing technology in a way that it is not requiring all of the user's attention. While it's clear from Don Norman's Principles of Interaction Design that feedback is necessary in order to make it clear to the user that an action has been taken and what has been accomplished, it is important to remember that there are different types of feedback, in addition to audio, including visual and tactile. Therefore, the Assistant does not always need to speak. In the case of the Google Home Assistant, the lights turning on is confirmation enough. Other ways to provide feedback to the user is through positive and negative tones, and light. Additionally, we can also add earcons, or distinctive sounds that are used to represent a specific event or convey other information (think of the sound you hear when you power up your computer), to this list.

Case Study

Metrics to define a successful user experience in the field of voice interaction and what can we learn from skills currently on the market

To shape the first draft of an evaluation framework for VUI, we randomly selected some VUI Skills available on the market with the aim of highlighting the most user-relevant VUI's properties (this exercise has been done without any commercial interest or purpose)

9 metrics for evaluating voice experiences

To evaluate how a VUI is performing, we can use a 5-Point Likert Scale (from -2 "I completely disagree" to +2 "I completely agree") to understand how much people agree or disagree with the following statements:

PERSONALIZATION	The VUI tailors and individualizes the entire interaction I have with it through the use of my user data and user preferences
CONVERSATIONAL	The VUI carries on the conversation beyond one turn. I feel that the conversation is natural, flexible
CONTEXTUAL	The VUI has an awareness of my physical and emotional state, and can refer back to my previous interactions with it.
SPEED	It is faster for me to use the VUI over the alternative(s)
NATURAL	I am able to speak to the VUI just as I would another human being. I don't have to remember what to say to the VUI or how to say it
COOPERATIVE	From my point of view, the VUI's turns are brief and optimally relevant. I don't feel that the system is talking away at me
RELATIONSHIP BUILDING	I feel that the VUI establishes a personal relationship with me: it is friendly and I can trust it.
DISCOVERABLE	The VUI helps me to understand what it is capable of. I find the VUI is easy to navigate, and I have a frictionless experience
CONVERSATION REPAIR	I feel that the VUI is able to handle errors by providing me with relevant solutions rather than with, "Sorry, I don't know that one"

LEGO® DUPLO® Stories

LEGO® DUPLO® Stories is an Alexa Skill targeted towards toddlers, which allows them to experience the joy of learning through short, interactive "Choose Your Own Adventure" stories involving animal play and vehicle play, using the LEGO DUPLO bricks they already own. If a child does not have all of the pieces, they can use whatever they have on hand to bring the story to life.

SCORING

		I completely disagree				I completely agree
PERSONALIZATION	1	○	○	○	●	○
CONVERSATIONAL	2	○	○	○	○	●
CONTEXTUAL	1	○	○	○	●	○
SPEED	N/A	○	○	○	○	○
NATURAL	2	○	○	○	○	●
COOPERATIVE	0	○	○	●	○	○
RELATIONSHIP BUILDING	0	○	○	●	○	○
DISCOVERABLE	0	○	○	●	○	○
CONVERSATION REPAIR	-1	○	●	○	○	○

COMMENT

Some negative comments towards the LEGO® DUPLO® Stories Skill include the frustration felt with regards to the skill's lack of ability to understand a variety of utterances, and its lack of ability to handle errors. As one user states, "[The skill] couldn't understand what we said; [it] kept jumping back to [the] start of stories, our son was disappointed and confused."

The positive comments towards the LEGO® DUPLO® Stories Alexa Skill reflect the personalization of the skill; children have the power to choose their own story path. Moreover, the skill pays clear attention to the environment within which children are using it—indoor play. It is clear that the LEGO® DUPLO® Stories Alexa Skill drives value for toddlers, with one parent stating, "The kids had fun with this story building; it let them play longer with the blocks then they normally would have."

Employing aspects of storytelling can help in the creation of actions and skills. By using storytelling, which allows us to present information in a way that resonates on an emotional level, we can build a relationship with the user. We can also use storytelling to take chaotic bundles of information and communicate them in a more coherent way, where using a story arc gives direction leading from one idea to the next.

Domino's Pizza

Domino's Pizza is an Alexa Skill that allows users to place and track pizza orders from Domino's Pizza. Users can either place their "easy order" or their last order (saved within their Pizza Profile).

SCORING

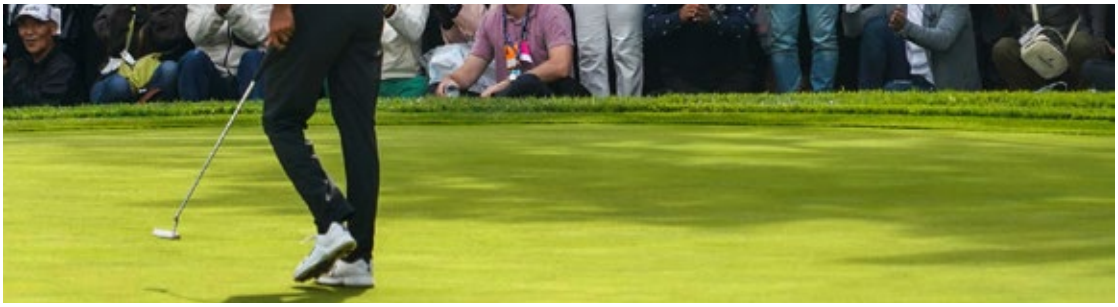
		I completely disagree				I completely agree
PERSONALIZATION	-1	○	●	○	○	○
CONVERSATIONAL	-1	○	●	○	○	○
CONTEXTUAL	-2	●	○	○	○	○
SPEED	2	○	○	○	○	●
NATURAL	-1	○	●	○	○	○
COOPERATIVE	2	○	○	○	○	●
RELATIONSHIP BUILDING	-2	●	○	○	○	○
DISCOVERABLE	-1	○	●	○	○	○
CONVERSATION REPAIR	-2	●	○	○	○	○

COMMENT

Some common user opinions reflect the desire for the Domino's Pizza Skill to be more proactive. Once the user has placed their order, they can ask Alexa for Domino's Tracker® updates at each stage. However, users express a wish for the skill to provide these updates without having to ask for them. One user states, "[I] would like [Alexa] to update with notifications as the order progresses." Another user states, "[The] only thing that would be useful to work better is, once you have ordered it, [Alexa] automatically lets you know it's in the car and on the way, instead of having to ask." It is clear that in order to drive value to users, we need to create our actions and skills to be proactive through guided conversation, engaged by contextual information.

The Domino's Pizza Skill struggles with error handling and getting conversations back on track. One user noted that when the Domino's Pizza Skill encountered an error, Alexa only responded with, "Pickled pepperoni, there is a problem." It was also clear from user's comments that the skill was not entirely thought through, and was more pushed out quickly with the intention to create hype. For example, the creators did not consider the case of applying coupons, which is something that the app alternative currently handles.

PGA Tour



PGA Tour is an Alexa Skill that allows users to keep up to date on the latest scores from every PGA TOUR event throughout the season. They can ask Alexa for the latest leaderboard or get the score of their favorite player.

SCORING

		I completely disagree				I completely agree
PERSONALIZATION	0	○	○	●	○	○
CONVERSATIONAL	1	○	○	○	●	○
CONTEXTUAL	2	○	○	○	○	●
SPEED	2	○	○	○	○	●
NATURAL	-1	○	●	○	○	○
COOPERATIVE	2	○	○	○	○	●
RELATIONSHIP BUILDING	-2	●	○	○	○	○
DISCOVERABLE	1	○	○	○	●	○
CONVERSATION REPAIR	1	○	○	○	●	○

COMMENT

From user opinions, it is clear that users cannot speak to the PGA Tour skill naturally; this is because the system struggles to understand a variety of utterances. As one user states, “[I] asked, “How is Jason Day doing today?”,” and the response they got was, “Sorry I can’t find the answer for the question I was asked today.”

On the positive side, the PGA Tour skill demonstrates an ability to handle errors and anticipate the conversation beyond one turn. As one user states, “If I ask it for the leaderboard and there is no tournament, it will tell me the name of the next tournament and when it starts.” In addition, the system exemplifies successful discoverability, meaning the VUI helps the user to navigate its capabilities. One user explains, “I really like that it asks you questions after it answers your query. After I ask for the leaderboard it asks me if I want the score for an individual player.” Moreover, the system represents how to create value for users by following Grice’s Maxim of Relation, and providing only the most relevant information to the user. As one user explains, “I love the ability to only get info on the few players that I am following.”Learning

Conclusions from Case Studies

By exploring the above case studies, the current capabilities and limitations of the voice assistant become apparent. While it is possible for assistants to employ some aspects of human-to-human conversation, in terms of asking questions and giving suggestions, they still grapple with applying context and remembering users. This is something that will come in time as Natural Language Understanding technology continues to develop.

Additionally, many voice experiences continue to struggle with handling errors, and accounting for the variety of utterances that users have to try and access the capabilities of the VUI. As we have previously mentioned, these are two key elements, that when taken into consideration in the design of an action or skill, can build user trust. Designers can help to set brands up for success by anticipating every way a conversation may go and reflecting that in their VUI designs.

Tools to Design VUIs

Tools to make voice and multimodal interface designs tangible for quick and dirty prototyping

Voiceflow

Great for designing pure voice interactions. Voiceflow is a tool for designing, prototyping, and publishing voice apps for Amazon Alexa and Google Assistant, with no coding needed. Voiceflow gives users the ability to visually design, and quickly iterate their conversational flow with a drag and drop Lego style block system, integrate voice apps with external APIs, and publish voice apps to both Amazon Alexa and Google Assistant.

CREATOR:

Braden Ream, Tyler Han, Michael Hood and Andrew Lawrence

PROS:

- Design for both Google and Alexa (project is agnostic to any platform until uploaded)
- Links to Amazon Developer & Google Developer accounts, and when a project is uploaded, generates an interaction model in the native console (Alexa Skills Kit SDK or the Actions console)
- Ability to upload and test with users on real devices within minutes (no coding required), as well as ability to make quick iterations
- Variables give you the ability to capture and store information in order to build better, more powerful, more personalized Alexa skills

Adobe XD

Great for designing interactions for second-generation smart speakers that have screens. Wireframe, design, prototype, present, and share amazing experiences for web, mobile, voice, and more — all in one app. XD is made for designers, creative teams, and organizations that need to design at scale. It's everything you need to work more efficiently, collaborate effortlessly, and create and manage your design system.

CREATOR

Adobe

PROS:

- Ability to design and prototype multimodal (voice and visual) interfaces

CONS:

- Adobe XD does not provide a lot of functionality for designing pure voice interfaces
- Adobe XD is not as valuable for testing as it does not allow for variations of utterances (the different ways a user might express an intent), and does not provide logic (the ability to respond to different choices and use variables)

Voice Apps

Voice Apps allows users to build and publish skills for Amazon Alexa devices (echo, dot, show, etc.).

CREATOR

Jeff Bolton

PROS:

- Voice Apps provides slot supporting; by using variables, which are placeholders for information, Voice Apps allows you to capture and store information to build better, more powerful, more personalized Alexa skills

CONS

- Not as intuitive as other software; collapses what the user would say and how the assistant would respond into one block
- To attach audio, it must be in a URL format, you cannot directly upload audio files

Botsociety

Botsociety allows users to preview and prototype voice assistants for the Google Assistant, Amazon Alexa and more.

CREATOR

Botsociety

PROS

- Unlike other prototyping tools, Botsociety automatically builds a map of the interface, essentially you just need to write dialog

CONS

- There is no ability to control the process of testing. The user just watches a video of the test, instead of interacting with the interface with their personal voice commands

NOTE

Botsociety is currently working on enhancing the capabilities, such as adding testing, of their Voice User Interface design and prototyping tool with a new platform, Botsociety Voice. You can request early access here: https://botsociety.io/voice

Methodology overview

The process for creating valuable voice experiences and the role of each design phase

Voice experiences are great when they offer a faster, easier, or more delightful way of doing things. As we have concluded in background research, it is important to create voice experiences that don't just create hype, but that actually add value to people's lives by marrying functionality with an enriching experience.

At Connect Reply we have developed a design methodology for crafting consistent and customer-centered experiences. Here, we apply the 7 design phases to the creation of valuable voice experiences.

DISCOVERY

To explore whether using Voice is the best option, over other channels, for the objective at hand by eliciting, clarifying and collecting business requirements and related KPIs. A proposal outline is produced to ensure the marriage of client goals with real value

UNDERSTAND & ANALIZE

To dig into current business processes, as well as the current experience for users with existing touchpoints, by mapping their experience. The aim is to understand users and their needs, as well as the technology we will be using to help meet those needs. Through this phase we will identify "aspirational" user requirements, and use cases where voice can benefit users.

CX DESIGN

To translate user requirements and key use cases into Sample Dialog, outlining the end-to-end conversation between the user and the VUI. In this phase we will build a persona for the Voice Assistant, and begin to explore possible responses people might have to the voice assistant.

UX DESIGN

To elaborate on Sample Dialog by using Detailed Conversation Design to account for all possible scenarios, or all the different paths that can be taken through the VUI system. Prototyping the solution to ensure that user and solution requirements are being met. The aim is to study, document and represent all the possible user experiences to determine whether exceptions and specific needs are viable to include in the conversation design.

UI DESIGN

To apply copywriting techniques in order to "dress-up" voice interactions, and to communicate the VUI's Persona, ensuring alignment with brand's tone of voice. In this phase we study interactions to ensure they are contextually relevant, and allow users to speak naturally.

IMPLEMENTATION SUPPORT

To guarantee implementation quality and adherence with the original design, as well as figure out implementation limitations that have an impact on the final product's CX, UX and UI.

MEASURE

Perform usability testing of the live product with real users, in order to measure performance and user satisfaction. The aim is to spot any problems and suggest improvements for future releases. We will use analytics to understand all the ways users engage with the VUI: how, when and why, to help us make iterations to better meet their needs.

References

- Aaron Marcus (Ed.). (2015). Proceedings, Part III, from HCI International 2015: The 4th International Conference on Design, User Experience and Usability. Springer.
- Pearl, Cathy. Designing Voice User Interfaces: Principles of Conversational Experiences. Sebastopol, O'Reilly Media, Inc., 2016.
- https://www.androidauthority.com/what-is-google-duplex-869476/
- https://arxiv.org/pdf/1901.06525.pdf
- https://blog.google/inside-google/alphabet/this-years-founders-letter/
- https://blog.google/outreach-initiatives/accessibility/impaired-speech-recognition/
- https://www.cbinsights.com/research/facebook-amazon-microsoft-google-apple-voice/
- https://chatbotslife.com/what-is-conversation-design-4cfe7ed200ea
- https://chatbotsmagazine.com/context-is-king-creating-intelligent-conversations-for-chatbots-9372bd6c2ca4
- https://codelabs.developers.google.com/codelabs/actions-2/#0
- https://design.google/library/conversation-design-speaking-same-language/
- https://designguidelines.withgoogle.com/conversation/conversational-components/errors.html#
- https://designguidelines.withgoogle.com/conversation/conversation-design/what-is-conversation-design.html#
- https://designguidelines.withgoogle.com/conversation/conversation-design-process/is-conversation-the-right-fit.html#
- https://developer.amazon.com/docs/alexa-design/get-started.html
- https://developer.amazon.com/docs/ask-overviews/requirements-to-build-a-skill.html
- https://developers.google.com/actions/assistant-investments
- https://www.drift.com/blog/conversation-design/
- https://fortune.com/longform/amazon-google-apple-voice-recognition/
- https://www.ibm.com/watson/advantage-reports/future-of-artificial-intelligence/ai-conversation.html
- https://www.interaction-design.org/literature/article/how-to-design-voice-user-interfaces
- https://medium.com/@blog_bleistift/voice-ui-use-cases-and-ux-challenges-29044471cf61
- https://medium.com/@edenspiekermann/step-outside-your-ux-bubble-86b090cdc2fc
- https://medium.com/@sachinrekhi/don-normans-principles-of-interaction-design-51025a2c0f33
- https://medium.com/@takashimokobe/the-evolution-of-conversation-design-chatbots-voice-user-interfaces-and-google-duplex-815d7bee2233
- https://medium.com/microsoft-design/voice-user-interface-design-new-solutions-to-old-problems-baa36a64b3e4
- https://www.mindtheproduct.com/2017/11/calm-technology-can-help-us-human-amber-case/?fbclid=IwAR0JcggqDbcTD8CiyY_XNdscH0sy1Ci0yLTpOrJq_5Pj4LhLdTV_hBXQqRA
- https://www.nngroup.com/articles/voice-assistant-attitudes/
- https://www.smarthomefocus.com/alexa-turn-on-lights/
- https://www.smashingmagazine.com/2019/05/future-design-voice-prototypes/
- https://theblog.adobe.com/91-of-brands-are-investing-in-voice-how-to-make-it-work/
- https://www.theverge.com/2017/7/10/15947672/amazon-alexa-voice-controls-shopping-prime-echo-how-to
- https://www.toptal.com/designers/ui/designing-a-vui
- https://towardsdatascience.com/conversational-ai-design-build-a-contextual-ai-assistant-61c73780d10
- https://voicebot.ai/2017/09/20/voice-app-retention-doubled-9-months-according-voicelabs-data/
- https://voicebot.ai/2019/04/15/smart-speaker-installed-base-to-surpass-200-million-in-2019-grow-to-500-million-in-2023-canalys/
- https://www.youtube.com/watch?v=O2_A0ViX2Pg
- https://www.youtube.com/watch?v=ZtnhzZ23P_E&t=707s

All links are updated to 31/11/2019