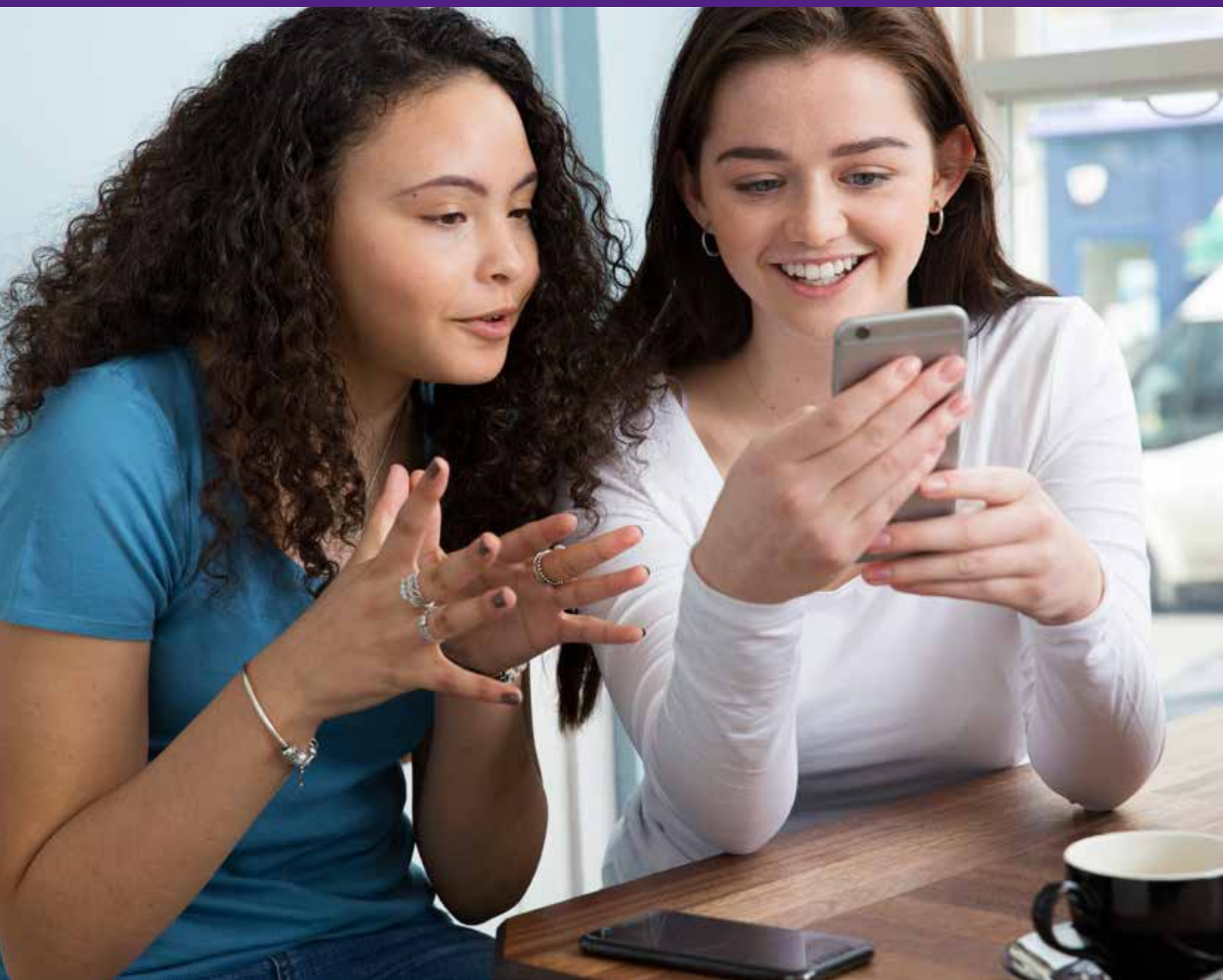
Talking to technology: will we be speaking to our gadgets in a post-PC world?

By Dr Nicola J. Millard
Head of Customer Insight and Futures



'The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it.'

Weiser (1991) [1]

In the next few years, we'll increasingly be talking to our technology, our houses, our cars. After all, speech is the way we talk to each other. So it seems reasonable to expect that it'll become the key way we talk to our technologies, too.

We could be entering a post-PC era as we wave goodbye to the mouse and the keyboard. The smartphone is our window on the world today, along with various voice-driven devices – from smart homes to smart cars to smart speakers. None of these lend themselves very well to keyboards – and that's where speech steps in. Gartner estimates that by 2020, 30 per cent of our interactions with technology will be 'conversations' with smart machines [2].

The idea of conversational user interfaces has been around for a while. But despite much recent success in natural language processing, communication between human and machine is still in its infancy. Now companies have access to larger and larger data sets – as well as processing power in the cloud – they can build voice technologies to have more meaningful interactions. (Although most aren't scintillating conversationalists just yet.)

Customers' attitudes towards talking to their technologies are also changing

69%
use Google Assistant, Siri or Cortana on their phone [3]

75%
use the same voice assistants on their smart speakers [3]

37%
of smartphone owners use voice technology at least once a month [4]

18%
use it at least weekly [4]

On the other hand, 23 per cent say they've only used voice technology once or twice. To me, that suggests some people find it frustrating or simply not useful. But only 12 per cent say they'd never use it.

So that's speaking to our phones and gadgets. What about speaking to companies?

Our Chat, Tap, Talk research with Cisco [5] found that 28 per cent of customers thought a voice-based chatbot would be an effective way to interact with a company (versus 37 per cent text based). They're one of the more visible examples of conversational technology – albeit generally messaging, rather than speech.

73 per cent of customers thought chatbots would improve their customer experience, particularly for simple things like getting train times, submitting meter readings and checking in for a flight. This doesn't mean humans are cut out of the loop, though.

74 per cent believed that there needed to be human 'checks and balances' in these responses – meaning there should be an easy route to the contact centre from a conversation with a bot.

Of course, the other issue is that we'll have so many interfaces across different devices. They need to be able to talk to each other, rather than interfere with each other. Longer term, chatbots and voice tech will probably start to merge, and we'll get a single 'digital butler' across everything.

The art of conversation

In the 1950s, Alan Turing proposed the 'Turing Test'. The name of the game? To fool a human into thinking they were talking to another human when, in fact, they were talking with a machine. Since then, we've spent many decades trying to hone the way we talk to technology – from basic conversational agents such as ELIZA in the 1960s [6] to natural language IVR, and now to chatbots and intelligent personal assistants.

Despite all this tinkering, despite all our technological advances and despite the explosion in digital data for machine learning to feed on – nothing has properly passed the Turing Test. So far.

So a new challenge has been set

The 'Winograd Schema Challenge' has become the new holy grail for speech technologies.

It's similar to the Turing Test but uses a series of set multiple choice questions that need common sense reasoning and an understanding of the real world to successfully answer. It can be a minefield, especially in the English language. For example, what's 'a French teacher'? A person who teaches French or French person who is a teacher? It all depends on emphasis and context.

And unfortunately, those are things machines haven't learned yet. The most infamous example is "call me an ambulance": your voice assistant will probably call you by a new name, not call for the emergency services.

None of this is helped by the ambiguity of human language. Saying you've been waiting "forever" doesn't actually mean that – it just means you've been waiting for a long time and are probably a bit annoyed. This instinctive intelligence, the ability to fill in gaps, and a lack of codifiable 'common sense' makes conversation a difficult technological nut to crack.

Language is one of the world's biggest and most tangled data sets

Just because I can understand French, doesn't mean I can have an effective conversation in French. For a proper conversation to take place, the machine needs to be able to answer back – conversation is a two-way affair. This calls for a surprisingly complex set of skills including speech recognition, speech synthesis, semantic analysis, syntactic analysis, sentiment analysis, common sense, real-world understanding and real-world knowledge. They all have to work together to create what we know as conversation.

It's not all doom and gloom, though

Natural Language Processing (NLP) is one area where technology has made giant leaps recently – error rates are now reportedly at around 5 per cent, which is about the same as a human. Partially this is down to better noise cancelling and microphones.

One of the biggest advances in conversational technologies has been a result of a different approach to deep learning. Rather than trying to figure out the exact meaning of sometimes-chaotic speech on the first try, or just repeating back what has been said to them, the deep learning algorithms which power these technologies work by trial and error. They draw on huge amounts of real human dialogue, generate a number of possible responses and rank them in terms of how well they match actual human responses. Then they can rapidly fine-tune provisional guesses.

The more popular voice assistants have the advantage of being able to learn from millions of conversations. But for progress to be made, they also need to learn from all their mistakes – to cumulatively learn from all the 'call me an ambulance' style blips.

The mistakes, however, are sometimes wildly inappropriate – which can cause risk-averse brands to play it safe. Often, they either hard code responses (rather than use machine learning), or use humans to monitor their responses for appropriateness before they go out to customers (more augmented intelligence than artificial).

The nirvana of 'ask me anything' is still a long way off

Machines trained to do a narrow range of tasks can perform surprisingly well, but their abilities tend to be biased towards the strengths of their manufacturers. Amazon's Alexa leans toward retail and entertainment. Google is strong on search. Siri, Bixby and Cortana are largely input devices for the hardware that they sit on.

All of these are typically very focused on specific tasks, like buying airline tickets, or relaying the weather forecast. By knowing a bit more about user context – like their location, who they associate with, what their diary looks like, what their routines are – they can go further. For example, telling you how the weather will be at home when you get back from work, or knowing who your partner is and texting them to say you're running late.

Companies are now putting more effort into 'chit-chat'

It's focused less on goals and more on the social aspects of conversation. Typically, these models need an extremely large data set to be trained – some are using dialogue from movie scripts, in fact. They also need knowledge in the area being discussed, and a 'memory' to hold the disparate parts of a conversation together. For example, if you ask "who's Jennifer Lawrence?" and then "how old is she?" the system needs to know those questions are connected and answer accordingly.

The way voice assistants are designed also calls for a different approach to those typically used to design voice systems such as natural language IVR. Up until now, dialogue design has largely been based on single, task-based commands.

In that sort of system, you'd plan a trip like this:

"Do you want to plan a trip?"	<YES>
"Say which city you'd like to go to"	<ROME>
"Did you say 'Rome'?"	<YES>
"Say when you'd like to go"	<SEPTEMBER>
"Say what date in September"	<17TH>

Frustrated yet?

Voice assistants change this because the focus is on conversation. So it would be more like:

"Let's plan this trip. What did you have in mind?"
<I'VE HEARD ROME IN SEPTEMBER IS NICE>

"Rome's a great choice. The weather is usually good there at that time of year. Did you have a specific date in mind?"
<AROUND THE 17TH>

Humans also tend to structure conversations in certain ways

Normally to suggest turn-taking and direction. For example, ending a sentence with a question is a natural cue for the other person to answer.

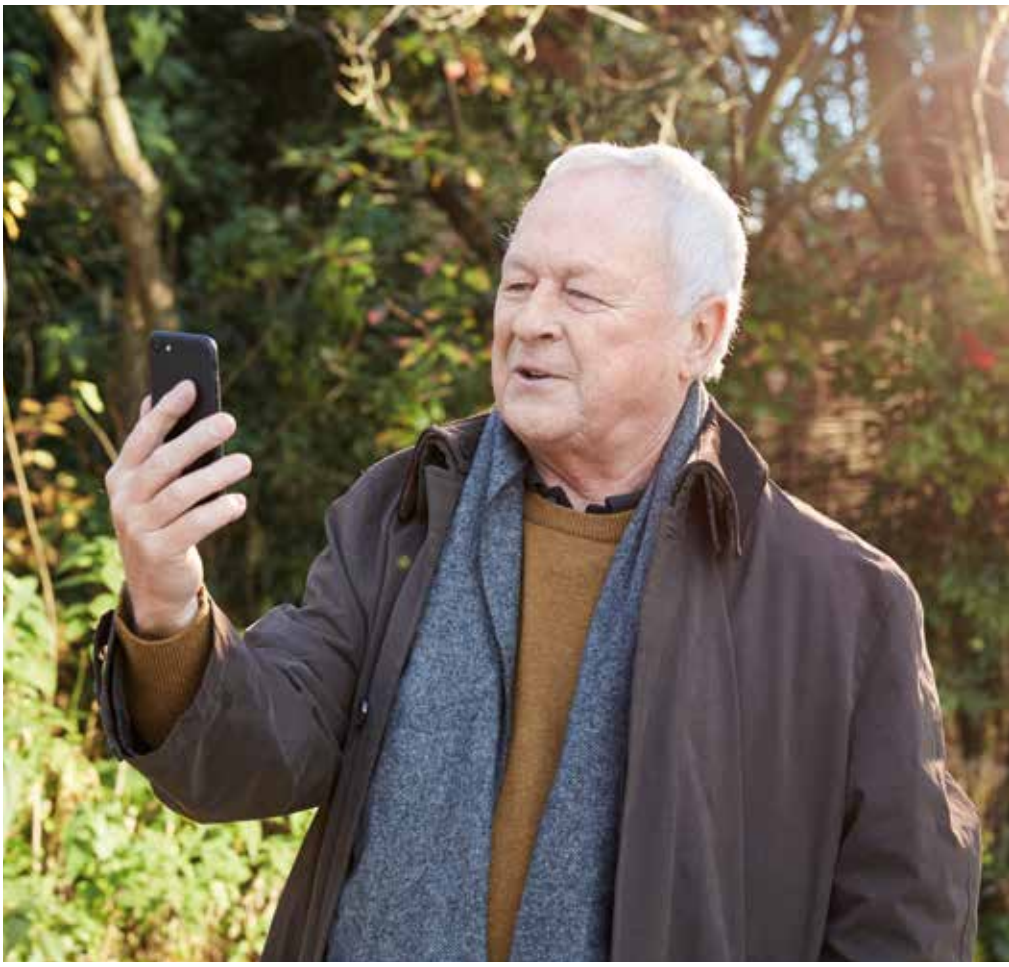
According to Amazon's Alexa guidelines [7], giving closed (rather than open) questions can get a better response. For example, "we have that in blue or red, which would you like?" rather than "what colour would you like that in?" Saying "thanks'" or "got it" shows the user's answer's been registered. In longer interactions, cues like "firstly", "finally" and "then" give users clues to how much longer they need to listen.

To create a better conversational style, many tech giants have hired comedians, poets, playwrights, journalists, screenwriters and novelists to script the interaction. One scriptwriter even compares writing for a voice assistant to writing an absurdist play [8]. That's because normal scripts try to develop character or advance the plot. Scripting a voice assistant is about developing a non-human persona who can get the user to their goal. The conversation is a means to an end and a 'happy path' exists to do this effectively – meaning that there are also a number of unhappy paths which don't.

Many voice designers spend a lot of time thinking about these 'unhappy paths'. By understanding the fail points, designers can figure out how to turn failures into success. The link to absurdist writing is that the human element is unpredictable and the voice assistant's role is to figure out how to steer them back onto that 'happy' path (assuming it can figure out what that is).

Do these systems need to pretend to be human?

We tend to interact with the world in human terms – through conversation and communication. As a result, we very easily anthropomorphise our technology. But should we bestow characteristics, or a personality, onto these virtual agents?



In an effort to create some kind of personality, many of the personal assistants have human names – Alexa, Siri, Cortana. Bizarrely, many of these have female personas (many commentators have attacked this as yet another example of Silicon Valley sexism). Google have notably bucked the trend, needing people to say "OK, Google" to activate their assistant. Maybe this is a good move: it dissociates the technology from a human persona and continuously reinforces their brand.

Believing a voice assistant is human could be dangerous territory

The recent demonstration of Duplex, Google's new, prototype voice assistant, caused a lot of debate in this area. Duplex made basic reservations over the phone but mimicked realistic human speech down to pauses, "um"s and exclamations. This caused a backlash around the ethicality of pretending that a machine is human. Google have addressed these concerns by saying that further developments of Duplex will make it obvious to the person on the end of the line that they are talking to a machine.

Aside from ethical considerations, this is probably a good move because pretending that the technology is human means that it's easy to believe that we can say anything to it and it'll understand. But as we've already explored, lots of things get in the way of that: unclear expectations, lack of long term memory, no understanding of customer intent and lack of guidance. Users can end up frustrated, which is probably what's behind the high levels of drop off in the current set of personal voice assistants – particularly when you hit the "I don't know", or "I don't understand" dead ends.

None of this adds up to human standard of conversation, or a particularly good customer experience. People don't necessarily need to have an enthralling tête-à-tête, but they do need things to work.

Perhaps we're the ones who need to adapt

In the past, we quickly learned the language of keyboards, mouse, swipes, pinches and taps. To make voice work, we might need to adapt our conversation as well.

If we know we're talking to a machine, perhaps we should adopt 'computer speak' – we know we can't use the same language we use to talk to humans. We don't generally have to think too carefully which words we use or how to phrase things when we're talking naturally – we just speak.

Talking to a machine is a bit different. We need to start the conversation with a signal of intention (remember "OK Google"?) which is then followed by the machine's response. This isn't an entirely natural conversational situation. Like many new activities, the cognitive load (how hard our brain has to work to adapt) may be high – and there may also be a period of frustration as the machine learns how we want to speak to it.

The risk here is that if the effort's too high, users will quickly disengage and the technology will gather dust in a corner.

Speak easy

One of the big factors driving voice technology is that it's fast and easy to use – especially for specific tasks. According to Forrester [9], 73 per cent of routine phone interactions are just a quick glance to check notifications, see the time or text. These types of interactions are ideal for voice, because the information's in smaller, consumable pieces.

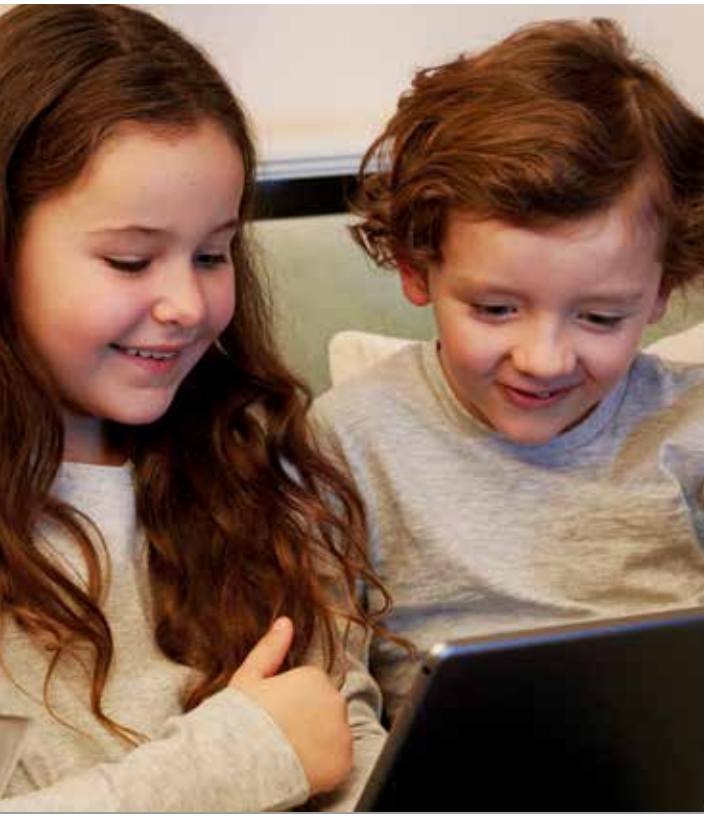
We can type 40 words a minute – and speak 150

It's easy for users to say "Play 'Bloodstream' by Ed Sheeran" instead of scrolling up and down a playlist. Equally, during an audio conference, it's easier to say 'record this session' than fiddle with buttons and keys. The added bonus is that simple tasks are emotion neutral and don't typically call for much sophistication in terms of language – "turn on the lights" or "turn up the temperature" are very specific commands that don't need much understanding. But if your command becomes more complex – like "turn on the lights on the top landing" – it might be easier to flick a switch.

Perhaps the biggest advantage of voice interfaces is that you don't need to be able to read, write, or have any manual dexterity to use them. This opens up new possibilities for many disadvantaged groups, as well as for young children. Much of the technology can be adapted for people with vocal impediments – so if they can't say "Alexa", it can be reprogrammed to trigger using another phrase or name.

It can turn children into masters of technology very early in life

As Metz points out in the MIT Technology Review [10]: 'already they're making an incredible amount of data and computer aided capabilities available directly to children – even those not yet in kindergarten – for learning, playing, and communicating. With Alexa, kids can get answers to all kinds of questions (both serious and silly), hear stories, play games, control apps, and turn on the lights even if they can't yet reach a wall switch'.



The downside is that they can very easily add their favourite sweets to the family shopping list without their parents noticing.

And they're particularly useful for exclusively hands-free environments like driving

Primarily, that's for safety. Distraction can be fatal at the wheel – tuning the radio, programming the sat nav and dialling the phone can significantly impact the ways we drive. A number of studies have found voice-based interfaces help to reduce distraction and improve vehicle control, speed control and lane keeping.

The other advantage for the car is that the vocabulary involved is limited – there are only so many things you want to say to your car – meaning responses are likely to be more accurate.

There are issues around designing in-vehicle voice systems, though. It's sometimes difficult for drivers to remember specific commands, so don't use the system so much. And bringing in a screen to talk to the driver when voice isn't working can distract them from the road.

So what's getting in the way?

Voice technology is often confined to private spaces. 22 per cent of people in the J. Walter Thompson and Mindshare Future study [4] said they'd feel embarrassed using voice interfaces in public. The same percentage also said they'd be wary about other people overhearing what they were saying (especially in sensitive transactions like banking). So it may take a few years before chatting to Siri or Cortana on a train, in the middle of the street, or at our desks becomes socially acceptable.

There are other issues hampering voice technology from being as easy as it could be. They include ambient noise (although many now use noise cancelling technologies, which help considerably), multiple voices, or when a system gets accidentally triggered by a common keyword – even on the radio or television. Without the advantage of facial expressions, body language, and gestures, technology can easily get the wrong end of the stick.

The inevitable "I'm sorry, I don't understand that" dead end is also frustrating for users. MIT's Cynthia Breazeal [10] suggests one potential solution would be to design the technology so it explains why it doesn't understand what you're asking, so users can reframe their question. This may make a lot of sense, but could get annoying fast.

Another fix for the dead end is to connect to other channels, including a real human. This ability – linking multiple channels together – is important to users. You might have ordered something on an app, but want to check its delivery status through voice.

We found customers get frustrated if they can't switch channels when they use chat and social media [5]. Voice interfaces aren't immune to this. They need to be able to connect through to a human (probably on the phone (potentially increasing voice demand into your contact centre), but text chat is another option). Some conversational applications already allow this. The potential difficulty is how much of the previous conversation can be passed through to the contact centre – customers don't like repeating themselves.

The other factor affecting how customers use this technology is the scope of what they're trying to do. Voice technology alone doesn't handle complexity well; it isn't a particularly information-rich environment. So if customers start by asking for something that doesn't make sense to the system, or which isn't possible, the whole interaction's doomed. Voice tends to cope better with simple, unambiguous tasks (like asking about one specific product), rather than vague ones (like finding out more about products in a certain category).

But a picture paints a thousand words

It seems more and more likely that companies will start to integrate screens with their voice technology. Just as we speak faster than we type, we read faster than we hear. Having a machine read a list aloud is far more tedious than glancing at it, and can test the limits of our short-term memory – we can't generally remember more than five options at a time. If there's only one 'top' result on voice, it works well, but multiple results can be problematic.

If we do introduce a screen into the equation, the voice interface needs to complement the visuals, not just repeat what's on screen. We can use screens to guide and structure next steps for the user; meaning technology can support the kind of complex interactions that voice technology alone can't quite handle.

Stalker or butler?

The most effective intelligent technologies aggregate data from multiple sources – calendars, to-do lists, email, browsing history – and create a personalised view of everything that matters to the user.

The best of them can send reminders for upcoming appointments, suggest travel plans based on that appointment, check for delays from traffic information, send weather reports and help people buy things. They can become our personal butler – organising us and anticipating our every need.

Problem is, this is all very personal information

It has the potential to cross into creepy territory – there's a fine line between a butler and a stalker. These technologies all have underlying algorithms that constantly learn more and more about us to incrementally improve the ways they interact with us. The extent to which the user feels this is a: benign and b: beneficial will determine whether or not they stick with it.

For example, if your digital butler knows you're approaching a pharmacy, would you want it to remind you to buy cream for that condition in a loud voice? Control needs to stay with the customer, and they need to give their permission as to which data the machine can learn from, and which they shouldn't.

In an era of GDPR, the things that could hold voice assistants back are regulation and customer buy-in. If you put the names of any of the major voice assistants into a search engine, it won't take long to find people talking about security and privacy.

> 44 per cent of people in the J. Walter Thompson and Mindshare Future [4] research said they were worried companies were listening to their conversations.

One major issue is the way we activate many of these devices. Unlike Star Trek, where you just needed to touch your badge to engage the computer, these technologies continuously scan their environment for mentions of their name. In other words, they're always listening. Although they don't link up with the cloud technology that powers them until they're activated, people are concerned these machines are eavesdropping on us – could hackers hijack them as listening devices?

Just as we've grown used to seeing a red light on a camera when it's on, it needs to be obvious when these devices are in listening mode – Alexa already does this by lighting up blue. And some people have suggested devices should alert anyone new who walks into a room that it's listening – but that might well become as annoyingly interruptive as "so-and-so has just joined the call" in an audio conference.

Of course, most brands want to be the ones to own this digital butler. After all, it's the ultimate way to get closer to customers: to personalise their relationship with them, to gather data on them and become more proactive for them. But the interesting debate in the future will be who owns this data? How confident are customers in giving permission to their digital butler to track their preferences and keep their data safe?

Branding and advertising might also become part of the voice battleground

Lots of brands might feel that they're losing their relationship with the customer, so they'll look to a branded voice assistant to show their personality. The problem then is that the actual voice it is delivered in may be created by the manufacturer of the device, not the brand. And critically, everyone wants the ability to grab the valuable data being gathered on customers. If brands create their own voice assistants the customer experience could get more complicated and fragmented – "I'm sorry, Alexa will need to transfer you now" – when really people are calling for everything to be as easy as possible.

Interestingly, voice technology might have a big role to play in the future of advertising too – an industry that used to rely on television screens and billboards to attract attention. As people increasingly turn to smart speakers, watches and cars – non-visual formats – traditional marketing methods simply won't work.

Just as advertisers purchase promotions on social media, they may also trigger targeted, paid voice promotions when users search for certain products. They might also sponsor certain conversations if particular keywords come up. This has the potential to seriously anger consumers if their easy and efficient service gets frequently interrupted by ads.

And it's all especially problematic if people feel they're being manipulated to buy certain products without their knowledge, or that the conversation's more useful to the owner of the technology than it is for them. Of course, voice technology might begin to operate on the 'Freemium' model, where customers pay a subscription to remain ad free. Or we could see a move to commission: when the customer acts on their recommendation, the company gets a fee – that's how many aggregation websites work at the moment.

To speak, or not to speak, that is the question

Voice technology is clearly a powerful move for the future of technology and customer experience. But it's still early days yet – we shouldn't overlook its challenges.

So if you're considering creating a voice assistant, here are the six things you need to think about.

1. Words, what are they good for?

How could this technology help your customers interact with you easily and effectively? How will you help customers get to their goal through natural conversation, without needing to remember complex keywords or talk like a robot? How do you show customers what they can and can't do?

2. It's all about easy interactions

Don't put automation in for automation's sake – it probably won't work. Figure out your company's most simple, repetitive transactions: that's what voice interfaces are ideal for. It's also where you can really polish up your customer experience; our research [5] showed customers saw advantages to using conversational interfaces to do quick and simple tasks like checking opening hours or train times, submitting meter readings, booking restaurants, or checking onto flights. However, if the technology is forced outside its narrow field of focus, the experience can quickly become frustrating.

3. Effective conversations can be difficult to achieve

Ironically, the one thing that chatbots (voice or text) are bad at is chat. Understanding words is one thing, understanding context and meaning is something else. They also have a very limited sense of 'memory' to help them understand when a conversation is connected, and when context has switched.

Real humans are better at chat, so look at how and when it is appropriate to involve your teams. If your customers hit a "I'm sorry I don't understand" dead end, how do you bring in the contact centre? Connecting conversations across multiple channels (thinking in an 'omni-channel' way) is just as important here as it is with chat or social media. How do you triage and route people to the appropriate agent, on the appropriate channel, smoothly and without your customer needing to repeat themselves? Will this result in more calls into your contact centre?

4. Security is key

Lots of people think voice devices are creepy because they continuously monitor your environment for keywords. Consumers are hot on privacy and security right now, so whether or not they properly take up these devices will depend on a trade-off between privacy, trust, security and potential benefits.

5. It's a branded experience

How do you bring your own personality to this technology when the manufacturer might own the voice and 'persona' your customers hear? (And potentially, the data that's gathered too.) How does it sit alongside the feel of your brand online, on the phone or in advertising? If you speak in different voices, you'll seem inconsistent, complicated and fragmented to customers.

6. A picture paints a thousand words

Voice assistants don't make screens irrelevant. Think about how you can use the two together to get the best out of both speaking and reading.

References:

[1] Weiser, M. (1991). The computer for the twenty-first century. Scientific American, September, pp. 94–110.

[2] Lownde, M. (2017), Innovation Insight for Conversational Commerce, Gartner white paper, 30 June.

[3] Accenture (2018), Time to Navigate the Super MyWay, Accenture white paper, https://www.accenture.com/t20180108T141652Z__w__/us-en/_acnmedia/PDF-69/Accenture-2018-Digital-Consumer-Survey-Findings.pdf

[4] SpeakEasy (2017), J. Walter Thompson Innovation Group London & Mindshare Future white paper.

[5] Davies, J. & Hickman, M. (2017), Chat, Tap, Talk: Eight key trends to transform your digital customer experience, BT/Cisco white paper, https://www.globalservices.bt.com/uk/en/point-of-view/chat-tap-talk-transform-your-digital-customer-experience

[6] Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. Communications of the ACM, 9(1), 36-45.

[7] Amazon (2018), Alexa Voice Design Guide, https://developer.amazon.com/designing-for-voice/

[8] Lin, M. (2018), Absurdist Dialogue with Siri, The Paris Review, 12th February, https://www.theparisreview.org/blog/2018/02/12/absurdist-dialogues-siri/

[9] Wise, J. Truog, D. Zoia, G. & Birrell, R. (2017), Q&A: Why Emerging Technologies Require Interaction Design Answers For CX Pros Who Need To Help Execs Understand IxD's Growing Importance, Forrester Research, August 3.

[10] Metz, R. (2017), Growing up with Alexa, MIT Technology Review, Vol. 120, no.5.

July 2018