EDISON

Mobile Ecosystems: Voice assistants

Its master's voice

Sector report, March 2017

Published by Edison Investment Research

Reality Bytes

Its master's voice

Voice is where much excitement and hype is to be found. Every ecosystem is developing a voice controlled digital assistant with which it hopes to enrich Digital Life and control the homes of its users. However, we find that voice suffers from significant limitations meaning that the user experience and functionality that it can offer is inferior to what can be achieved using a visual device. This combined with the fact that the intelligence of voice based systems remains rudimentary at best means that voice is unlikely to replace screens any time soon.

- **Three stages of understanding.** We define three stages in a machine's development to be able to understand voice commands: **1)** high word accuracy; **2)** understanding of the request in multiple word orders and formats; and **3)** understanding of context and circumstance. In our view it is not until machine understanding reaches stage 3 that voice can have any hope of challenging the established man machine interfaces of screen, touch, keyboard, haptics and mouse.

- **Defining voice.** Despite these limitations, voice usage in ecosystems is growing rapidly. Our research indicates that voice usage is growing only as an alternative to typing a request rather than as a rich two-way voice interaction with the ecosystem. Hence it is important to separate the two types of voice usage to understand voice's place in the Digital Lives of users. We term these as one-way voice and two-way voice.

- **One-way voice** is where voice is used as an alternative to using a keyboard. Most ecosystems have reached stage 1 making this use case viable. While input is voice based, the response is delivered through the usual visual method. Our research suggests that the vast majority of voice requests in digital ecosystems use this method which will have no effect on the monetisation methods currently used by Google, Facebook etc.

- **Two-way voice** is where voice is used as both input and output. It remains almost exclusively the realm of home speakers such as Amazon Echo and Google Home. We find that the rudimentary artificial intelligence (AI) of digital assistants combined with the limited amount of information that voice can convey, often has these systems falling back on displaying results on a screen.

- **Voice in ecosystems.** Digital assistants and voice-based interfaces are driven entirely by the AI that powers them. Although the search engines are leading the development of AI, all systems are far too rudimentary to replace visual based devices for the foreseeable future. Facebook is still the laggard compared to its peers when it comes to advances being made in AI.

Analyst

Richard Windsor +1 415 830 6032

rwindsor@edisongroup.com

Key contacts

Dan Ridsdale
Global Head, TMT
+44 (0)20 3077 5729
dridsdale@edisongroup.com

Rachel Carroll
Global Head, Investor Relations
+44 (0)20 3077 5711
rcarroll@edisongroup.com

William Bishop
Corporate Business Development, TMT
+44 (0)20 3681 2535
wbishop@edisongroup.com

Annabel Hewson
Institutional Equity Sales
+44 (0)20 3077 5700
ahewson@edisongroup.com

Peter Molloy
CEO, US
+1 646 653 7026
pmolloy@edisongroup.com

Klaus Schinkel
Director, Head of Germany
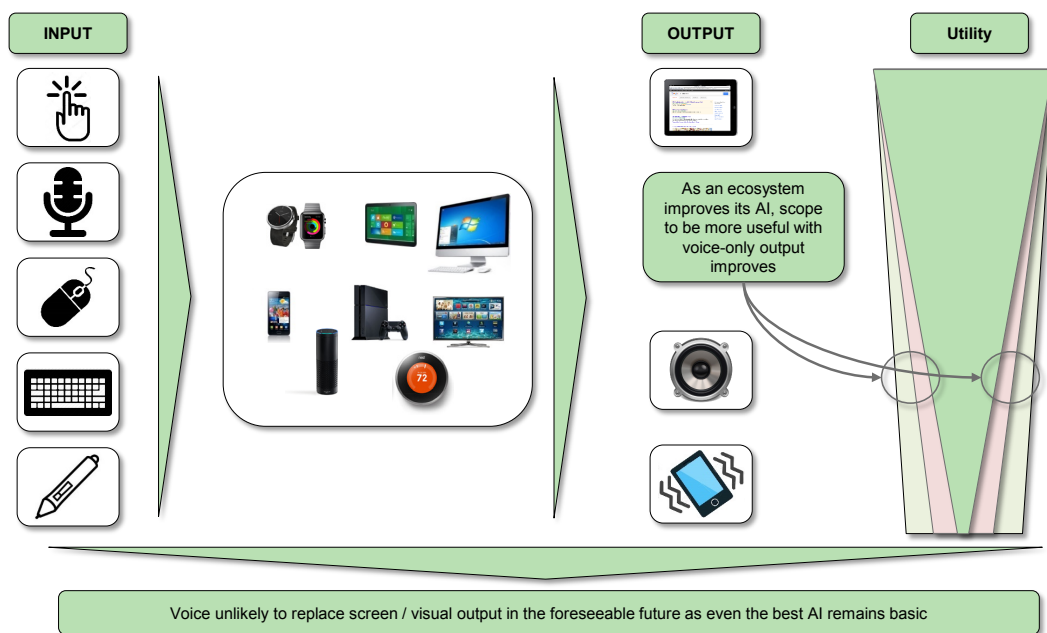+49 69 78 8076 960
kschinkel@edisongroup.com

Peter Bristowe
Founding Director
+61 (0) 2 8249 8343
pbristowe@edisongroup.com

Will Manuel
Director
+44 (0) 20 3734 1007
wmanuel@edisongroup.com

Voice as the user experience

Voice is currently an area of great interest in the ecosystem. At CES 2017, Amazon stole the show with virtually every developer of a smart home product or service enabling voice control with Amazon's home speaker, Echo. This combined with the hype around digital assistants and connected cars has meant that voice communication is an area of great interest. However, voice is not without limitations. In some circumstances it is an ideal method with which to communicate with a machine but for general use, its utility is still very limited when compared to using a screen and touch for communicating with the user (Exhibit 1).

Exhibit 1: Ins and outs of the man machine interface

Voice unlikely to replace screen / visual output in the foreseeable future as even the best AI remains basic

Source: Edison Investment Research

Improvements in voice recognition and natural language processing have meant that in terms of being able convert speech into text, the machines have recently become extremely accurate. Previously, machines would have difficulty in recognising the actual words spoken let alone derive any meaning from them. Hence, voice interaction with a machine was a slow and frustrating experience which meant that most users would try it and then revert to a keyboard. With much more accurate voice recognition, this is beginning to change as voice is now a reliable and user friendly method of text input into a machine. This is what we define as stage 1 of voice recognition in the ecosystem (Exhibit 2).

Exhibit 2: Stages of voice recognition in the ecosystem
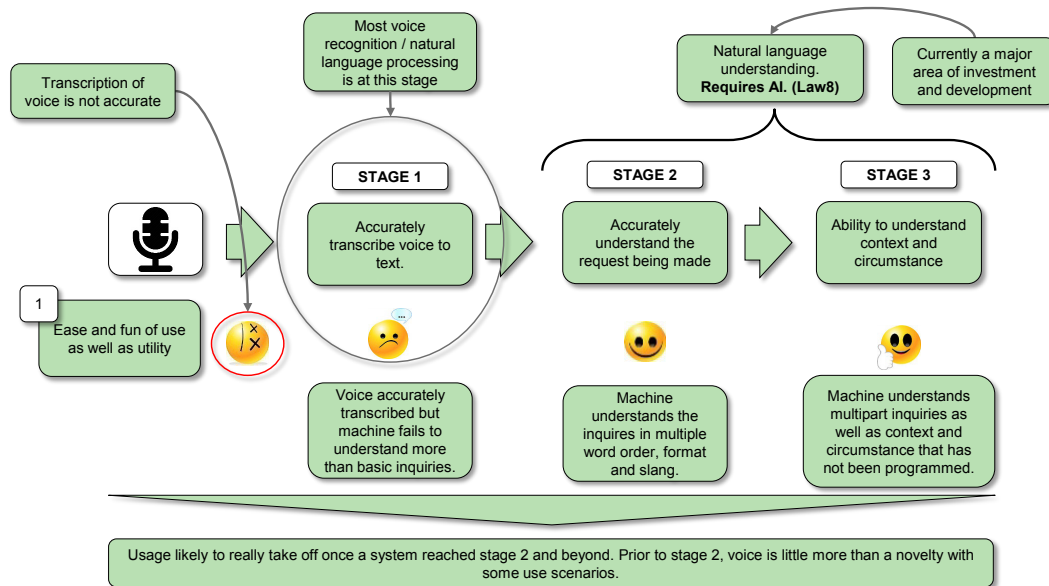
Source: Edison Investment Research

As a result of improved accuracy, usage of voice for data input has increased significantly where Google says that more than 20% of all Google's searches now begin with a voice request rather than a keyboard based input. However, the same is not true for the other side as almost all of the results of these searches are delivered on a screen rather than by voice. Consequently, we believe that significant progress is needed before a voice based system can rival traditional systems in terms of their ease of use and utility, if ever.

This is where AI comes in because to be really useful, the machine has to be able to understand and respond as if the user was talking to another human. Today this is very far from the case and even the best digital assistants are extremely limited in terms of what they understand even though they can accurately identify almost every word spoken. **This is what is referred to as natural language recognition which we think remains at a very early level of development (stage 2, Exhibit 2)**. Once this has been conquered, there is likely to be another upswing in usage but it is not until the machine can understand context, circumstance and multipart inquiries (stage 3, Exhibit 2) that voice as a means of receiving input from a machine for Digital Life services in the ecosystem will become a viable alternative to screens, touch, keyboards and mice.

It is at stages 2 and 3 that the difficulties really begin because both require sophisticated algorithms that are created using AI techniques. Our research has previously shown that these take a very long time to develop meaning that those who have been doing it the longest have the greatest advantage. (see Mobile Ecosystems – Men and boys, 24 February 2017). This is why progress beyond stage 1 has been quite slow despite rapid advances in improving recognition accuracy over the last few years.

The net result is that while users are quite prepared to use voice to input commands, they are much less willing to receive the results of their requests or inquiries solely over the voice channel (Exhibit 4) and tend to do so only when there is no other choice. This is why it is important to separate the use of voice in Digital Life services into two categories. We define and examine these categories below.

Two-way street

To fully grasp the position of voice in the delivery of the digital ecosystem, we must separate the two roles that voice plays. This is important as one (one-way voice) of these roles makes very little difference to the existing experience the user has with his digital device, while the other fundamentally alters it (two-way voice). Furthermore, two-way voice also places substantial constraints around the advertising-based monetisation that will require it to be completely rethought should two-way voice ever become the mainstream method by which a user interacts with digital devices.

One-way voice

One-way voice refers to using voice or audio based input exclusively for the input segment of transaction. In effect, it is simply a replacement for typing, with the response from the digital device being displayed in the same way that it has always been via a screen, audio or haptic feedback (Exhibit 3, top). This is often accompanied by voice recital of the results, but there is little doubt that the main medium used is visual. **In our view the vast majority of all voice based interaction currently fits into this category.**

Revenue generation from the advertising led monetisation model and subscription based model has always been driven by the output from the device rather than the input and consequently, it remains unchanged. Most importantly, the demands that are placed upon the quality of the voice based user experience are limited solely to voice recognition. It is when an ecosystem must construct an entire user experience around voice (two-way voice) that the problems begin.

Exhibit 3: One way voice vs. two-way voice
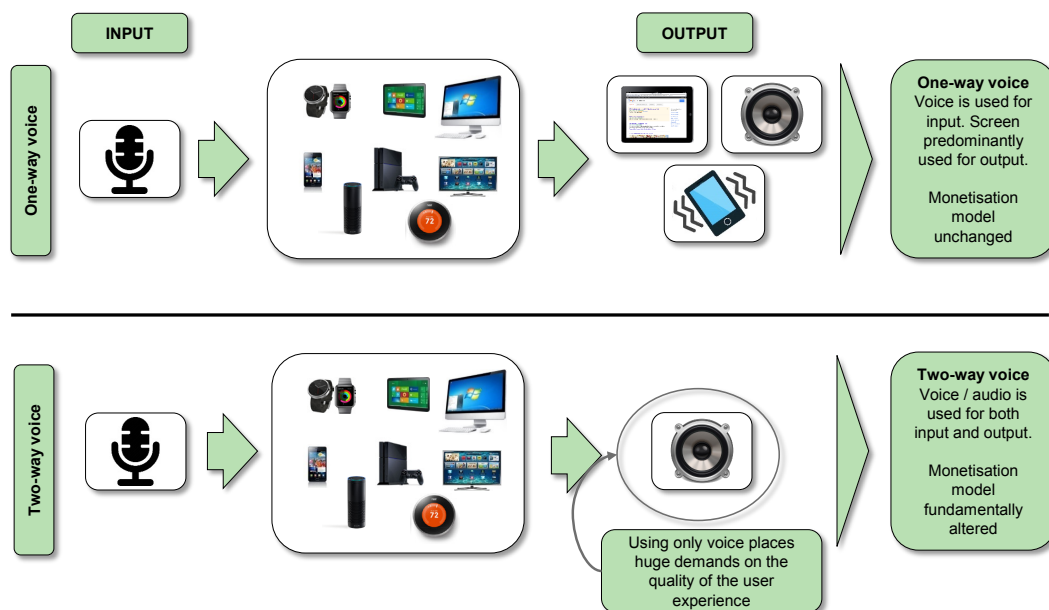
Source: Edison Investment Research

Two-way voice

Two-way voice refers to a use case where audio is the only method by which the digital device provides output in response to a voice command from the user (Exhibit 3, bottom). It is clear that two-way voice is a far more limited output method than using a combination of screen, audio and haptics. **This is because two-way voice is only utilising one (hearing) of the five senses in contrast to three (sight, hearing and touch)** that are made use of normally. It is this limitation that

has led to two-way voice only really being useful when the user has no other choice (Exhibit 4). In our view this is due to:

1) Voice-based services are not intelligent enough. Digital assistants are predominantly designed to interact with the user using speech such that they can more closely emulate the ease of communicating with another person. The idea is that at some point, talking to the assistant will be almost the same as speaking to a human which will be able to anticipate the users' needs as well as understand and solve multipart, complex problems. It is at this point that a number of Digital Life activities like search, browsing and shopping could, in their majority migrate to voice. However, the problem is that all digital assistants do not really understand natural speech and instead must be spoken to in some form of code. For example, in order to get Amazon Echo to play a playlist on Spotify a series of command words must be given in a specific order otherwise the desired music does not play. This is one area where Google Assistant is clearly ahead of its competitors as naturally asking for a playlist in series of word orders almost always produces the desired result.

Voice based services and digital assistants are the first real services to be offered to ecosystem users which are completely dependent on AI to offer a good quality service. This began with the ability to search the Internet using voice rather than text and has gradually evolved from there.

Consequently, almost all responses given by even the best digital assistants are effectively pre-programmed. For example, for a digital assistant to provide useful exchange rate data, it has to be taught at a very high level about currencies such that when it looks up data from a public source, it knows what to look for and how it should relay the data back to the user. What tends to happen when a request is made outside of that field is that the assistant transcribes the inquiry into a search box and displays the results on a screen. When the assistant is present in a speaker it will do this by alerting the user that the results of the inquiry have been displayed in the smartphone app that comes with the speaker. This is effectively a fall back to one way voice. Our research suggests that this happens far more often than it should for two-way voice to become acceptable as a reliable method for the man machine interface.

The more that AI can be advanced to make digital assistants more intelligent and intuitive, the less this will happen and the more that the assistant will be able to produce meaningful and useful results. We believe that this will be a key factor in determining which assistants get used (Exhibit 6), creating monetisation opportunities for their owners and which reside on the device gathering cyber dust.

2) Voice-based interaction is very limited in terms of the information that it can convey. In advertising, it is generally accepted that images can convey far more information to the user than a description. This idea has been present in journalism for over 100 years and has often been expressed by philosophers and teachers. Consequently, we believe that interacting with a machine using two-way voice puts substantial limitations on the level and richness of the interaction that can be achieved. This means that either the user experience must be extremely well designed with a highly intelligent agent or that the utility of such a system will be limited. **This is the main reason why we think that conversation-based user experiences will not replace visual systems, but instead will continue to be mostly used where other output options are not available or are not convenient to use.**

In our opinion, it is this limitation that explains why voice is only the preferred mechanism of interaction with a machine when it is difficult or impossible to use other forms (Exhibit 4). The most common examples of these are when hands are occupied in the home such as cooking or in the car where the manipulation of a device is both difficult and dangerous.

Exhibit 4: Reasons and places for the use of voice

Reasons for using voice output

Location where voice primarily used

Inability to use other senses is by far the biggest reason to use voice.

61%
30%
24%
22%
12%
1%

43%
36%
19%
2%

Hands/vision occupied | Faster results | Easier than typing | Fun | Direct access | Other

Home | Car | Out and about | Other

Voice output is very far from being at a level of quality that would enable the replacement of visual systems as an output method.

Source: MindMeld, Edison Investment Research

Outside of these sorts of use cases, the limitations of voice are so detrimental to the user experience that the user is likely to always prefer to receive the output through a visual medium of some description. As the intelligence that underpins the voice interface improves, so will the user experience (fig. 1) but we do not think that it will not replace the visual/touch-based interaction system for the foreseeable future.

Voice in the ecosystem

Almost all of the major ecosystems are working on a digital assistant that provides voice based interaction with the user. This began with the ability to search the Internet using voice and has gradually evolved from there. Apple was indubitably the first (2011) to migrate this into a service but it is Google that we think has the real edge (Exhibit 5). The idea is that this will become as useful and dynamic as having a human assistant creating a reason for a user to choose one ecosystem over another.

Personal assistants are being delivered predominantly through two main device categories, the mobile phone and the home speaker. The home speaker opportunity was opened by Amazon with its Echo product but as this category lives and dies by the quality of the AI being offered, we believe that others have an opportunity to displace it. Top of this list is Google Home as it has by far the most advanced AI (Exhibit 6) that already offers users the best experience outside home automation.

Digital assistants are at an extremely early stage and are not even remotely close to approximating human intelligence or even outwitting a young child. Consequently, these assistants are designed to alleviate the user of the easy, humdrum tasks such as finding a route, affecting a search or creating reminders and calendar entries. The digital assistants are starting simple because to improve, they need to generate usage as all AI algorithms need vast amounts of data to evolve. This creates a difficult chicken and egg problem as a digital assistant needs to register a good score on Edison Laws of Robotics 1 and 2 (easy and fun user experience (Exhibit 5) to generate usage but at the same time will have difficulty in becoming useful without the data needed to train them. This is why the digital assistants are starting at a very basic level where in many cases they offer little more than voice control of apps or a straightforward search.

Exhibit 5: Digital assistants assessed against Laws 1 and 2
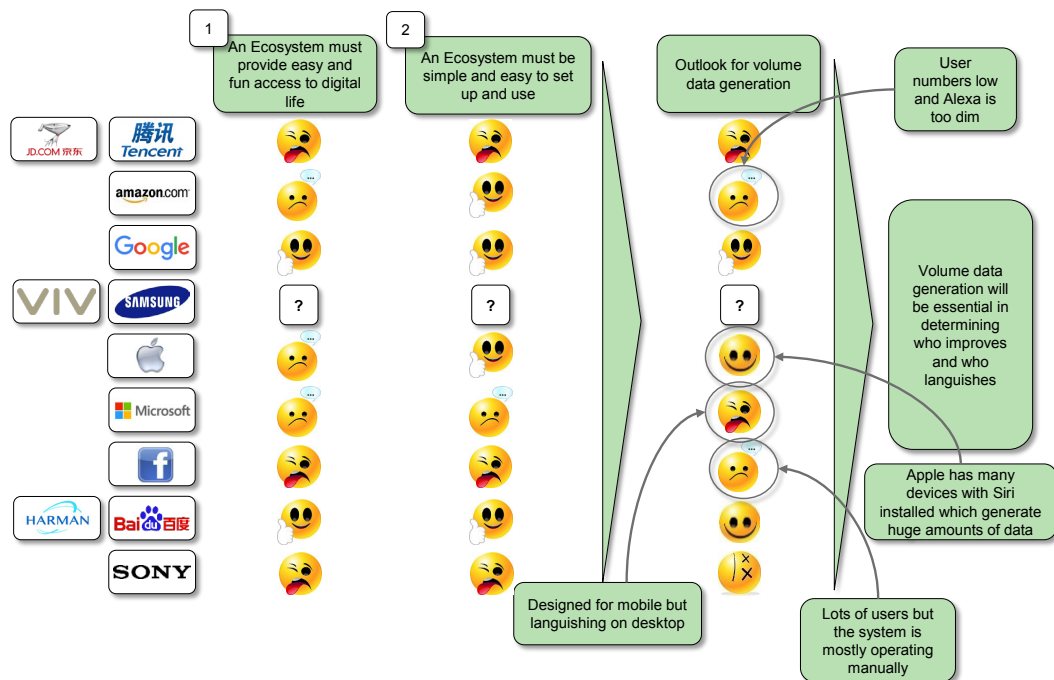
Source: Edison Investment Research

When it comes to data generation, we see Google, Apple and Baidu as having the biggest advantage. Google and Baidu have an edge because they already generate vast amounts of data with their search and in our view have the best AIs (Exhibits 5 and 6). Consequently, as they can offer the best user experience, they are likely to be used the most and therefore evolve and improve more quickly. Apple's Siri (figs. 5 and 6) also has an advantage for two reasons. **First,** it is present and set as default on the devices of over 400m users. **Second,** those devices generate far more traffic than an equivalent android device presenting Apple with a great opportunity to train Siri. To date, we think that data generation has not been very effective as we continue to see Siri on par with Cortana (Microsoft) and way behind Google.

Exhibit 6: Law 8: Artificial intelligence assessment by ecosystem
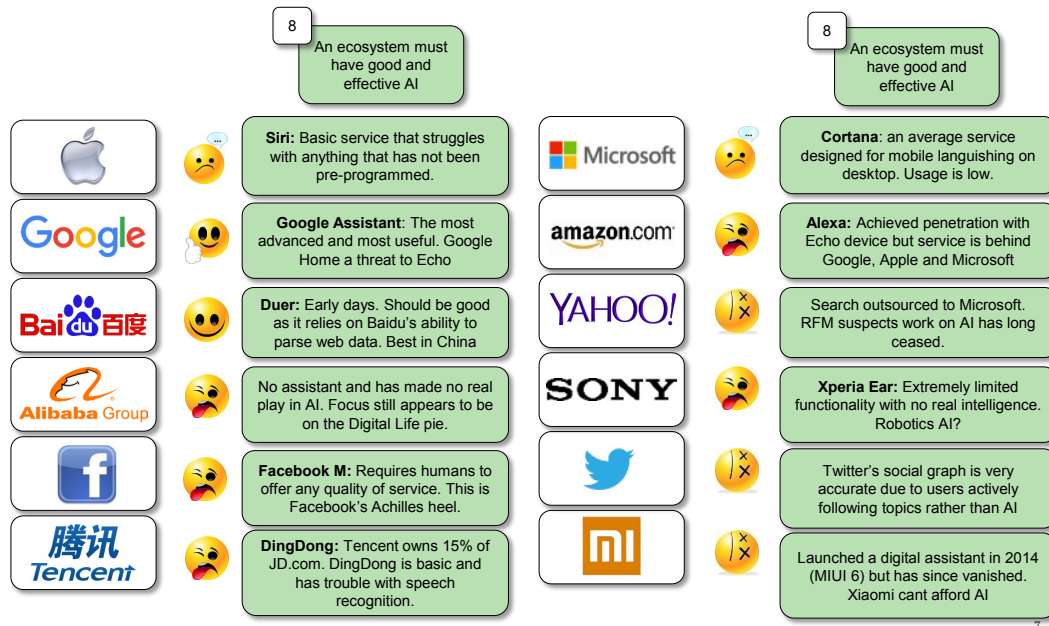
Source: Edison Investment Research

Conclusion

While the use of voice in the ecosystem is currently all the rage, it is really only seeing significant usage as a replacement for touch or keyboard based input. This is largely because voice is only capable of using one of the human senses to convey its message while the traditional systems are capable of using three. We believe that this means that the digital assistants that make up the voice agent will have to become far more intelligent before two-way voice can begin to compete with the utility of using a screen and a keyboard. This will be determined by the AI that drives the user experience of the assistant and here it is the search engines Google and Baidu that lead the field (Exhibit 6). This gives them the opportunity to create voice based user experiences that are far superior to everyone else's. The importance of intelligent voice is certainly rising and in some instances (home and car), it is the best form of communication, but we doubt that voice will replace screens at any time in the foreseeable future.

EDISON