STRATEGY ANALYTICS
Research, Experts, and Analytics

The Role of Voice in IoT Applications

Report Snapshot

As IoT applications proliferate they present new and innovative ways to collaborate, communicate and interact, both from a human and machine perspective. While most of the focus in the IoT has been on data, the integration of voice – including VoLTE – into IoT applications offers a versatile method to provide human interaction; communication and control, that with proper implementation and consideration for relevance, costs and human factors, can provide a more flexible user experience in a more economical way than traditional methods such as touch screens or data input.

This paper examines the previously under-explored role of voice in IoT and identifies use cases and opportunities where the inclusion of voice offers not only inherent value to the specific solution, but also makes practical economic sense.

December 2015

Andrew Brown
Tel: +44 1908 423630
Email: abrown@strategyanalytics.com

Contents

Exhibits

1. Executive Summary

Speech is a singularly efficient way for humans to express ideas and desires. It is therefore unsurprising that we have always wanted to extend this expeditious communications mechanism with and command machines by voice.

As IoT applications proliferate they present new and innovative ways to collaborate, communicate and interact, both from a human and machine perspective. While most of the focus in the IoT has been on data, the integration of voice into IoT applications offers a versatile method to provide human interaction; communication and control. And with proper implementation and consideration for relevance, costs and human factors, voice can provide a more flexible user experience in a more economical way than traditional methods such as touch screens or data input.

The growth in service providers implementing IMS, which enables voice over LTE (VoLTE), and the rapid roll out of LTE services in many markets, in addition to existing private VoIP/SIP networks, represents a new and exciting opportunity to cost effectively integrate voice into different IoT applications. Naturally, not all IoT applications will merit voice integration, but this paper outlines and showcases the multitude of applications that would benefit.

Strategy Analytics believes that voice has a pivotal role to play in the Industrial IoT segments on the shop floor where automation, robotics and mobility will factor in. Specifically, Strategy Analytics projects that voice could capture up to 12% of Industrial IoT applications by 2022 and be especially prevalent in the lucrative Asia-Pacific and North American regions.

In the consumer segment, voice will become especially pervasive in wearable IoT and the consumer healthcare segment, Smart Home, consumer electronics and Smart Buildings. In these vertical industries, voice has the potential to capture up to 18% of connections in the 2020 to 2022 timeframe.

There will be 3 key types of voice communication in IoT environments:

1. Bi-directional voice communication

2. Mono-directional voice communication

3. Voice Recognition

Strategy Analytics believes there are 4 key reasons that voice is suited to a range of IoT applications:

1. **Speech is the natural mode of communication for humans.** It is both intuitive and easier to convey commands verbally.

2. **Voice recognition is particularly appealing when the human's hands or eyes are otherwise occupied**. For example it may not only be convenient but also a legal requirement to use verbal commands (if any) rather than be distracted by typing or using other means of communication, as in distracted driving legislation, for example.

3. **Voice telephony is an efficient means of bi-directional voice communication with machines that can listen, and respond without the need for complex commands.**

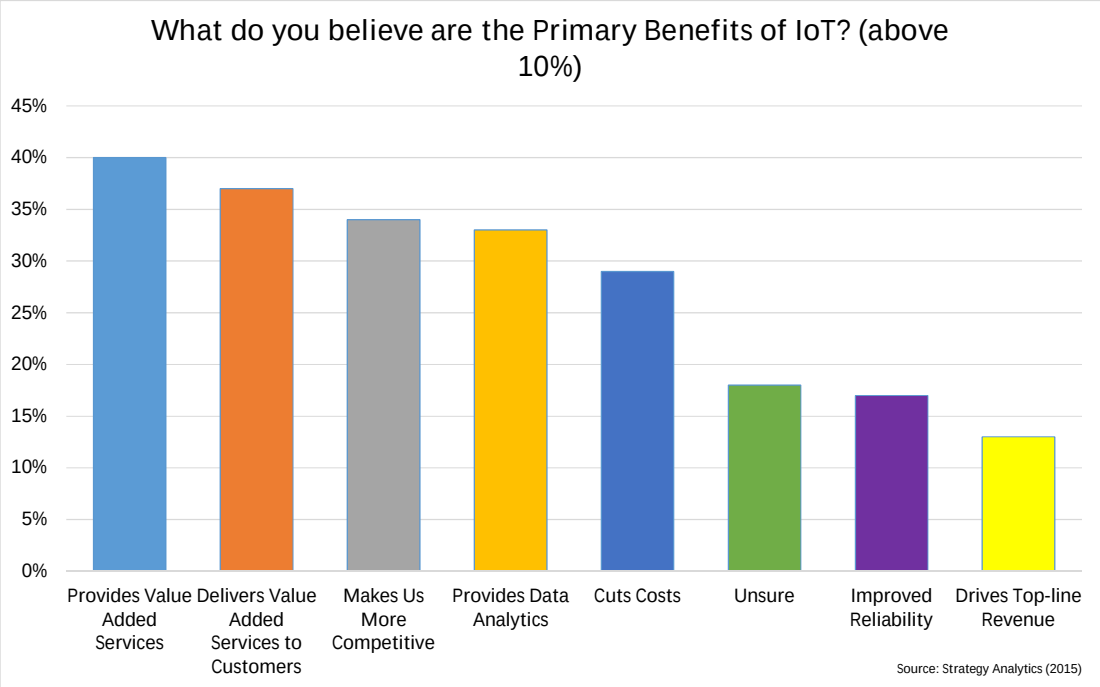
4. **Cost saving factors: Voice integration could potentially challenge the need for a touch screen on many devices,** as it reduces the cost for devices that will be dormant for the majority of the time. There is a point at which devices that are not used frequently (e.g. a few times a year), do not need a touchscreen. Voice would be more cost effective to simply offer connectivity to a call centre to achieve the same function. This represents a shift from a volume to a transactional model e.g., a million devices that generate 50,000 calls per year are arguably better serviced via a voice connection than deploying a $5 dollar screen on a million devices.

2. The IoT Opportunity is Significant and Growing

The Internet of Things will be the defining technology trend of the next decade and will impact virtually every vertical industry and permeate the vast majority of business processes. IoT is an integral component of a broader digital transformation. The primary objectives of IoT are: to strengthen customer engagements, in order to increase customer loyalty and accelerate market penetration, and to streamline operational efficiency to improve organizational effectiveness and profitability.

The growth in IoT platforms (PaaS), coupled with the declining costs of connectivity and hardware has fostered an environment where numerous sectors can now connect their assets to realize the benefits of IoT. As Strategy Analytics' primary research shows (Figure 1), drivers are primarily focused on those transformative digital elements such as driving new value-added services, becoming more customer-centric, offering competitive differentiation and better understanding their own environments and processes through Analytics.

Figure 1: What do you believe are the Primary Benefits of IoT? (n=450)



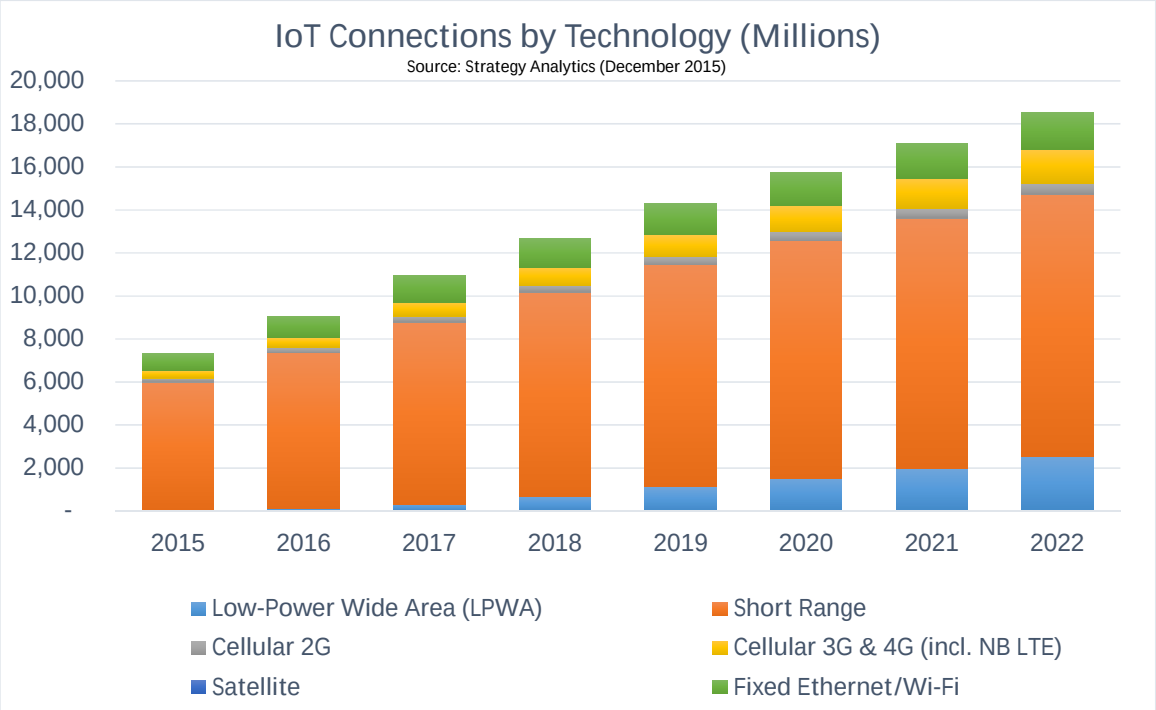
Source: Strategy Analytics 2015

Historically the focus has been on cutting costs or driving efficiency savings (the 5th most important component), but increasingly companies are seizing the opportunity to use embedded intelligence and connectivity to make the transition from selling products to selling services.

One example would be global Kärcher, a leading worldwide provider of cleaning technology for private, commercial and industrial deployments. Kärcher transitioned its business from selling products to selling cleaning services through the implementation of IoT that monitors products and fleets.[1]

One way of measuring the growth in the IoT is in the number of connections (as illustrated in Figure 2). In 2015 SA estimates there will be just over 7 billion IoT connections (excluding smartphones and tablets), which will grow to over 18.5 billion by 2022. The majority of these connections will come from short range devices such as intelligent, embedded sensors. This will enable greater instrumentation and facilitate deeper, contextual understanding of the functioning and performance of assets, than ever before.

Figure 2: IoT Connections by Technology Type



Source: Strategy Analytics 2015

[1] https://wcs.eop.vodafone.com/Vogeo/m2m/karcher-offers-its-customers-an-innovative-solution-for-the-optimisation-of-fleet-cleaning-solutions

3. IoT is all about the Data, Right?

The perceived assumption about the IoT and indeed M2M connectivity, is that the focus is solely on data. And while it is true that the data deluge is real and will continue to grow exponentially, data is not the ***only element*** that merits consideration in IoT environments. Voice, too, has a crucial role to play.

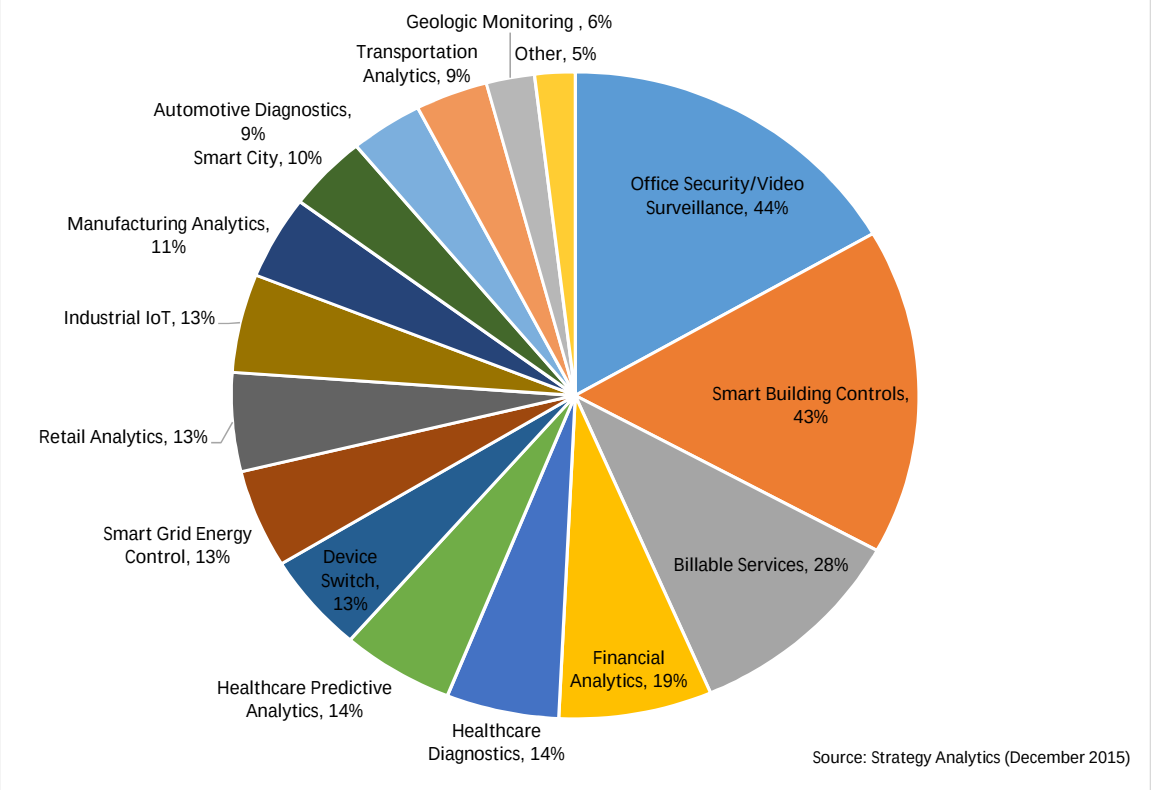
Massive amounts of data are created, processed and transmitted daily, as every movement, transaction and choice we make becomes digitized. Big Data Analytics is revolutionizing the way we live, work and do business.

The amount of data generated daily and annually is astounding. IBM estimates that 2.5 quintillion bytes of data is created every day and that 90% of the world's data has been created in the last two years. According to Cisco, although the number of IoT connections is growing threefold, global machine IP traffic will increase 15-fold between 2014 and 2019, from 308 petabytes in 2014 (0.5 percent of global IP traffic) to 4.6 Exabytes by 2019 (2.7 percent of global IP traffic).[2]

Additionally, an overwhelming amount of the data that we generate is unstructured: that is, it does not fit neatly into rows and columns, but rather it comes from a vast array of sources and applications ranging from text messages, Emails, photos and selfies that are posted on social media sites, digital audio and video, imaging files like X-rays and CT-scans, cell phone GPS signals, online purchasing transactions, satellite data etc.

[2] http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI_Hyperconnectivity_WP.html

Figure 3: Planned IoT Deployments by Types of Services and Solutions (n=450)



Source: Strategy Analytics 2015

In terms of applications being deployed, smart buildings, smart cities, healthcare, utilities, retail, industrial and manufacturing and vehicles all represent areas of investment as highlighted in Figure 3.

While it is true that data is integral to the IoT and indeed Big Data Analytics is a core element in the transition from M2M to IoT, one area that has been underexplored is the area of voice and its use in embedded applications.

3.1 The Changing Role of Voice

Despite the fact that voice has, for years, been the core component of the Telecom industry, the market focus has shifted to data. Fixed and mobile voice revenues for many carriers have continued to slide, even on personal devices such as smartphones, impacted by over-the-top (OTT) data-based applications offering VoIP telephony. By contrast data volumes continue to skyrocket.

Nevertheless, voice continues to play a vital role for many companies in terms of customer interaction with devices. For example, many larger companies offer natural language processing (NLP) recognition software and virtual assistant programs for their communication and entertainment products that talk to users and respond to a list of commands, such as Amazon FireTV, Amazon Dash, Microsoft Cortana, Google Now and Apple Siri. Some even monitor user habits, answer questions guided by input and deliver suggestions based on logs of what users view, access or save, such as movies, music and favorite restaurants. These companies are all competing fiercely in the area of voice recognition and monitoring on their platforms. The functions do not just relate to replacing keyed-in commands, but also areas such as search.

Companies such as Nuance Communications have a major role in providing and integrating server and embedded speech recognition engines into many consumer products, such as smart TVs, but there are many other applications for voice recognition that span a range of applications from consumer to industrial, where voice recognition can play a compelling role.

4. Making the Case for Voice in IoT

Speech is a singularly efficient way for humans to express ideas and desires. It is therefore unsurprising that we have always wanted to communicate with and command our machines by voice.

Prior to the Industrial Revolution and steam engines, power for agricultural machinery was supplied by Oxen or other domesticated animals. Humans discovered that the same animals that provided the power for the machine could also understand rudimentary voice commands that allowed for management and control of the animals and the process, for example, when to begin pulling a plough or when to stop. Simple vocabulary was sufficient for a single human to control the movement of a large farm machine.

Naturally voice commands were not the only means of controlling animal-powered machines. A system of more direct commands was also available through the reins attached to the animal. However, in many cases, voice commands offered clear advantages over the alternative as it was both natural and intuitive. And it left the human completely free to do other things, such as walking alongside the machinery and throwing crops into a wagon, for example. This was efficient as it eliminated the need for an extra person to drive the machine, and the convenience of not having to return to the machine to issue commands greatly improved the efficiency of the operation (reins were always vital to guarantee control, of course).

Clearly, this reliance on the modest intelligence of the animal source of power was severely limiting, and even that limited voice recognition capability disappeared as animal power was replaced by steam and fossil fuel power. However, the allure of voice interaction with machines has remained and became stronger as technology became more advanced and complex.

4.1 Advantages to voice-machine interaction

There are a number of obvious advantages to voice-machine interaction:

- **Speech is the natural mode of communication for humans.** It is both intuitive and easier to convey commands verbally.

- **Voice recognition is particularly appealing when the human's hands or eyes are otherwise occupied**. For example it may not only be convenient but also a legal requirement to use verbal commands (if any) rather than be distracted by typing or using other means of communication, as in distracted driving legislation, for example[3].

- **Voice telephony is an efficient means of bi-directional voice communication with machines that can listen, and respond without the need for complex commands.**

- **Cost saving factors: Voice integration could potentially challenge the need for a touch screen on many devices,** as it reduces the cost for devices that will be dormant for the majority of the time. There is a point at which devices that are not used frequently (e.g. a few times a year), do not need a touchscreen and it would be

[3] http://www.ghsa.org/html/stateinfo/laws/cellphone_laws.html

more cost effective to simply offer connectivity to a call centre to achieve the same function. This represents a departure from a volume to a transactional model e.g., a million devices that generate 50,000 calls per year are probably better serviced through a voice connection than putting a $5 dollar screen on a million devices. There are exceptions to this, such as the scenario where the device is not a basic terminal and the users have a smartphone or wearable for a screen.

- **Voice-enablement is a fundamental tool for those with disabilities for whom data entry is not a viable option.** Voice communication with machines or bi-directional communication with an operator can be used to perform mundane tasks such as open doors, and also to communicate well-being remotely. For example, to enable independent living, solutions have emerged that allow operators to detect trips or falls using sensors such as accelerometers. However, the addition of voice recognition can function as a vital back-up to these solutions. And they also serve to mitigate the risk of false alarms and can provide a valuable source of onsite comfort.

5. Use Cases for Voice Integration into IoT offerings

With a vast increase in the number of connected devices, it is inevitable that voice, given its intuitive means of interaction and suitability for a range of applications, will pervade multiple different segments. These are summarized in Table 1. The potential range of applications for voice in IoT are broad and range from consumer to business and industrial applications.

Voice is a key attribute in terms of simplifying user interaction, but failure to consider the current and future voice requirements for an application. It is critical is that with any voice implementation, the type of functionality required fits appropriately with the kind of function that is required and that the addition of voice is not only necessary, but does not add significant cost.

5.1 Consumer v Business and Industrial Applications

The IoT market is broad and encompasses a range of consumer, commercial and industrial applications where voice can play a role. There are significant differences between the drivers for implementing voice into consumer products and from those that drive the same technology in the consumer market. For instance, companies face a different set of challenges when they try to implement audio applications in building systems, vehicles or develop new health care products. Consumer products are mostly commodity devices, with price the primary market driver followed by product features. Voice is a feature that does not need to make any consideration for infrastructure, other than the need for an Internet connection, and can be included or omitted from different products SKUs. Little consideration is given for product lifespan, component selection or platform or software scalability.

5.2 Consumer Applications

Consumer applications for voice include virtual assistants on smartphones as well as devices that do not include integrated telephony functions, such as wearable devices with minimal screen real-estate. Devices in this category include smartwatches and fitness wearables that can offer hands-free voice activation of a multitude of functions, through to smart televisions and games consoles.

According to Nuance, Smart TVs are great example of the value of voice integration into products, where they have seen the number of Smart TV customers using their voices instead of remote controls climb by 230%, and the number of transactions grow by over 300%.

“In an unconnected TV, the remote control was perfect: pick a channel. When we got cable and hundreds of channels, the seams started to strain. Now that cable TV provides thousands of on-demand titles, and TVs are connected to the internet, the remote is looking a little Victorian. At Nuance, we believe that voice is the best user interface (UI) because it is the natural UI of people – we like saying what we want, how we want to[4].”

[4] Dan Faulkner: http://whatsnext.nuance.com/connected-living/voice-and-internet-of-things-connected-devices-ces-2015/

In the consumer segment, voice will become especially pervasive in wearable IoT and the consumer healthcare segment, Smart Home, consumer electronics and Smart Buildings. In these vertical industries, voice has the potential to capture up to 18% of connections in the 2020 to 2022 timeframe.

5.3 Commercial and Industrial IoT Applications

What works well in products with high volumes (e.g. game consoles or smart TVs) does not work well for commercial and industrial markets, where often there is a legacy telephony infrastructure, and voice integration cannot be managed with a point-solution approach. In business environments, for instance, voice has often become part of a hosted SIP solution. However, with the enormous investment in LTE deployments by mobile network operators, it is increasingly expected that IMS VoLTE will become a preferred mechanism for delivering voice even to these vertical market applications.

There are multiple traditional IoT use cases that could compel the use of voice services in these vertical markets. For example, there are thousands of voice solutions for Inventory control, Health Care Monitoring, Logistics and Delivery (UPS/FedEx), Rental Car Management as well as applications from previous bi-directional paging markets. Additionally, a number of new commercial and industrial IoT applications may benefit from including voice functionality. As these vertical markets look to the future, LTE is increasingly expected to be the cellular connectivity standard of choice, and within this trend we can expect that many of these applications will leverage the native VoLTE capability of these networks to provide valuable voice services.

Voice-enabled technology is particularly suited for use in robotics and automation in a number of Industrial applications:

- Manufacturing: Shop floor.
- Automotive: Robotics assembly.
- Unmanned and piloted vehicles: Robotics in everything from unmanned drones, aircraft, space, ground and underwater exploration submarines.
- Healthcare: To assist in robotics during surgery. While we have this capability now the possibilities if robotics operatives were able to communicate with surgeons to improve surgery precision are significant e.g. “You need to guide a little to the left?”
- Agriculture: To more precisely spray crops and fields and tell them when the job is complete.

There are multiple business use cases for implementing voice into a commercial IoT environment:

- **Improved customer service:** Integrating support directly into the equipment (manufacturing equipment, ATMs,)
- **Infrastructure consolidation and reduced installation costs**: For example running on a central data backplane (building systems)

- **Providing more flexible service delivery:** Seamlessly integrating multiple locations (nursing and health care facilities, security systems, customer service)

- **More flexible customer support options:** Centralizing customer service support and offering that support in multiple languages (e.g. retail)

Commercial and Industrial customers are generally looking for voice solutions that can be easily integrated with their applications. It is not practical to offer solutions with large up-front licensing costs and multiple complex licensing requirements. Because VoLTE is increasingly 'native' in network operator LTE deployments, these additional costs and complexities can be avoided.

Strategy Analytics believes that voice has a pivotal role to play in the Industrial IoT segments on the shop floor where automation, robotics and mobility will factor in. Specifically, Strategy Analytics projects that voice could capture up to 12% Industrial IoT applications by 2022 and be especially prevalent in the lucrative Asia-Pacific and North American regions.

6. Types of Voice Services in the IoT

In order to deliver voice services in IoT, it is important to consider the different use cases and voice requirements for various market segments.

Strategy Analytics has identified three main types of voice implementation:

1. Bi-directional voice

2. Mono-directional voice

3. Voice recognition

6.1 Bi-directional voice

Bi-directional voice in IoT is common in human to human interaction via a human machine interface (HMI). The rampant growth in network connectivity combined with an unabated demand for communication and content, has resulted in media over IP (internet protocol) becoming the biggest consumer of bandwidth globally. This demand has prompted IP network providers and equipment vendors to become media mature, implementing IMS infrastructure in order to meet customer demand and counter-act the threat to voice revenues from OTT (over-the-top) voice applications. Applications can take advantage of the benefits packet switching has, while projects such as webRTC, promise a whole new voice and video experience.

The bandwidth required for bi-directional voice is higher than simpler mono-directional voice interaction, but is now more viable given the IP network infrastructure offered by many carriers. The demand for bi-directional voice can be met in a number of ways: for example using VoIP or SIP, or via VoLTE (Voice over LTE, which is based on the IP Multimedia Subsystem (IMS) network, with specific profiles for control and media planes of voice service on LTE).

VoLTE will eventually be supported by carriers of LTE networks globally (at present at least 43 networks globally already support VoLTE with at least another 51 imminent). Additionally the cost of providing an integrated VoLTE offering is falling as LTE chipmakers also expand their partnerships to offer voice in embedded applications, even over narrowband implementations (e.g. LTE Cat 1 devices[5]). VoLTE adds negligible cost to a voice-enabled device that incorporates an LTE modem as its connectivity solution.

For example, in 2015, 4G chipmaker Sequans Communications and Ecrio developed a joint solution integrating Ecrio's operator-certified FlexIMS Voice over LTE (VoLTE) and SMS over IMS client software with Sequans' LTE chipset platforms, offering a turnkey solution to support VoLTE on LTE phones, tablets, CPEs, gateways, M2M/IoT devices[6].

[5] http://www.radio-electronics.com/info/cellulartelecomms/lte-long-term-evolution/ue-category-categories-classes.php
[6] http://www.sequans.com/press-release/sequans-ecrio-deliver-volte-solution/

The upshot of these market developments is that the cost of providing voice capability in embedded solutions is falling. There is no longer a need to add a costly circuit switched MSC (mixed-signal circuit) to a cellular IoT network implementation, provided the core network supports IMS, VoLTE is a viable option for IoT.

An IP based voice solution, like VoLTE, is more cost effective than adding circuit-switched voice to LTE single-mode IoT devices, and it will generally provide a better user experience. On the network side, many operators have deployed dedicated core networks for IoT. These networks are usually packet only and adding a circuit MSC (mixed-signal circuit) or connecting the dedicated packet core with legacy circuit core is more complex and costly than using IMS VoLTE.

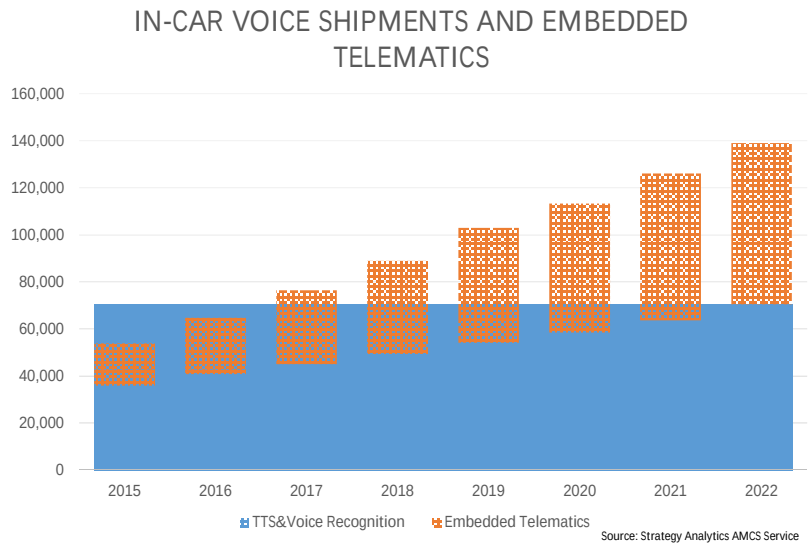
6.1.1 Bi-directional voice use cases

There are multiple use cases for bi-directional voice. Human voice communication to a contact center via a human-machine interface is a key function. This includes **security applications including:** built in IP video surveillance or voice integrated monitoring functionality (see, hear and broadcast) as well as panic button services (an intruder is in the building). Connection to a call center can be automated in these scenarios.

Voice integration is key in **vehicles**, both from a command and control standpoint as well as bi-directional voice services. OEMs continue to invest a great deal in voice recognition systems for in-vehicle infotainment systems. Speech has the potential to be the safest input modality, as it allows the driver to keep eyes on the road, and hands on the steering wheel at all times. Voice recognition systems are particularly well-suited for complex tasks that would typically require multiple steps when using a touchscreen or hardware controller. These tasks include SMS entry, destination entry, playback of a specific song or song subset from a brought-in device, and radio tuning. Other services such as breakdown services (or BCall) as well as concierge services (where is and how do I get to) are other functions in addition to legal requirements in many markets such as ECall[7] or stolen vehicle tracking in markets like Brazil[8].

[7] https://ec.europa.eu/digital-agenda/en/ecall-time-saved-lives-saved
[8] http://www.denatran.gov.br/resolucoes.htm

Figure 4: In-car voice and Telematics Forecast in Passenger vehicles



Source: Strategy Analytics

Table 1: IoT Voice Application Use Cases

Mode of Voice Application	IoT Market Segment	IoT Application Category
Bi-directional voice	Home Security	Security call centre/emergency services
		Voice-integrated video monitoring and broadcast
	Retail	Customer interaction
	Industrial	Training
		Maintenance/integration into manufactured equipment
	Healthcare	In-hospital care and outbound patient calls
		Wearable healthcare device integration
		Remote monitoring/Patient care e.g. trip or fall confirmation/assistance
		Remote interaction between patients and medical professionals
	Vehicles	Infotainment and other HMI functions
		Ecall and Bcall services
		Concierge services
		Rental cars
		Logistics/fleet management
		Stolen vehicle notifications/tracking
	Robotics	Manufacturing (e.g. robotic assembly), military (drones, recon), healthcare (e.g. surgery), (e.g. crop spraying)
	Wearables	Industrial and medical devices. Fitness trackers and smartwatches.
Mono-directional voice	Building Automation	Door entry controls/authentication via voice recognition
		Smoke/fire alarms
		Temperature monitoring (too hot or cold)
	Public Safety	PTT for emergency communications
	Robotics	Manufacturing (e.g. robotic assembly), military (drones, recon), healthcare (e.g. surgery), (e.g. crop spraying)
	Smart Cities	Public Transportation system notifications
		Public Address Systems
		Intercoms
		Music in shopping malls/elevators/retail etc.
		Automated announcements
		Mass notification in disaster scenarios/public emergency
Voice Recognition	Home Automation	Lighting Control
		HVAC/Thermostat
		Home Security Systems
	Home Appliances	Smart TVs/Home Theatre
		Games Consoles
		Other electrical appliances e.g. baby monitors
	Vehicles	Control and Command functions
	Consumer Devices	Smartphone, PC, wearable and tablet voice assistants

Source: Strategy Analytics 2015

Healthcare applications represent another area of opportunity for bi-directional voice implementation. Incorporating voice using VoIP or VoLTE makes sense, as a network connection with the patient can be maintained even when he or she is being moved around a hospital. Intercom push buttons for a direct connection to medical staff can connect to the same network. Voice could automatically be routed to different locations, or even different medical

staff to accommodate shift changes. In addition, voice could function as a regular telephone through a gateway to the public phone system, allowing patients to stay in touch with friends and family.

Other uses include remote in-home health care monitoring, especially support for the elderly. There are already healthcare devices that can detect trips or falls, but these do not always yield accurate results, relying instead on analytics to build a pattern of behavior with alerts confirmed by deviations in behavior patterns. Voice can support this system with the patient confirming whether they need assistance or if there is a simple false alarm. It can also be useful in emergencies if a voice channel is opened in the case of a trip or fall, or, for the patient to request help-similar to the automated response capabilities designed for ECall legislation in vehicles.

6.2 Mono-Directional Voice

A number of voice functions in IoT applications will simply require voice to be transmitted in one direction, or at most the transmission of one audio source to many receiving devices.

Although technically bi-directional, “point-to-point” communications could be enabled on the same network infrastructure using Push-to-Talk (PTT) or Push-over-Cellular (POC). Currently PTT/PoC is more of an enterprise service. It has been more of a North American phenomenon due to existence of iDEN and other PTT networks although there is increasing interest in enhancing PTT/PoC to meet the needs of US and European Public Safety for next generation Mission Critical PTT over LTE, as evidenced by the creation of the 3GPP SA6 working group.[9]

Full-duplex operation on mobile networks is made possible by separate frequencies for transmission and reception. Push-to-Talk services, offered by some mobile carriers, add functionality for individual half-duplex transmissions to be sent to another party on the system without needing an existing connection. Since the system is half-duplex, only one user can transmit by PTT at a time; the other party is unable to transmit until the transmitting user un-keys his or her PTT button.

The advantages of PTT are, of course, that intercom-type “point-to-point” functionality could be provided over the same network as tow-way communications, as well as meeting the needs of the public safety vertical.

6.2.1 Mono-directional voice use cases

There are multiple use cases for mono-directional voice communications in IoT. While it is true that many use cases may already implement one way voice broadcast, most of these involve human actors over public-address systems or recordings. It is possible to automate these broadcasts in real time with voice synthesis to provide real value in certain scenarios, or eliminate the issue of proximity, to minimize real-estate costs e.g. centralize operations for security monitoring/CCTV in other cities, rather than needing to localize these services.

Most mono-directional voice operations are viable in public spaces, especially around smart city scenarios, for example transport notifications (the train or bus is on time or late, for example).

[9] http://www.3gpp.org/specifications-groups/sa-plenary/sa6-mission-critical-applications

Others examples include:

- Public Transportation system notifications
- Intercoms
- Music in shopping malls/elevators/retail etc.
- Automated announcements
- Mass notification in disaster scenarios/public emergency

Mono-directional voice is particularly useful in broadcast for one-to-many scenarios and the growth in IP voice networks, including both VoIP and VoLTE as well as PTT/PoC on LTE networks, will bring down the cost of integrating broadcast or point-to-point voice in many relevant IoT applications.

6.3 Voice Recognition

Voice recognition allows a range of functions to be controlled by means of voice on a number of different device types such as computer operating systems, commercial software for computers, mobile devices (smartphones, tablets), cars, call centers, and internet search engines such as Google. By removing the need to use physical input mechanisms (or at least offering another, more intuitive option), users can easily operate appliances with their hands full or while doing other tasks.

Voice recognition systems are speaker-independent, and can respond to multiple voices and languages. They are also capable of responding to several commands at once, separating vocal messages, and providing appropriate feedback, accurately imitating a natural conversation. In order to achieve voice recognition functions, implementations must be either native via a host processor, or remotely by using a client-server approach. Speech recognition requires a certain level of system resources (CPU / RAM), so therefore native implementations are not suitable IoT applications with a small footprint using microcontrollers.

The use of voice commands are most popular on smart devices and search engines and use hosted (server-based) implementations even though the device resources are capable of processing speaker independent speech. The end-point is responsible for detecting and encoding the voice input, packaging the data and transmitting it to the server. The server is responsible for processing the audio data, decoding speech and performing any post processing required to improve context dependency and overall quality/confidence.

Analytics is becoming an important element in the IoT and the information captured by popular voice-recognition engines, such Apple Siri or Microsoft Cortana, are extremely valuable to those companies in terms of understanding their users, as well as the kind of commands they are issuing, which, when analyzed, can provide significant insight that may assist in future product development as well as feature sets for development in areas such as home automation (e.g. Apple Homekit).

Hosting voice recognition offerings also drives the need for connected products, subsequently driving data revenue for service providers, as well as helping vendors improve quality of the speech engines at the back end where the audio data is processed, the speech decoded and post-processing undertaken.

6.3.1 Voice Recognition Use Cases

The most common examples of voice recognition systems are on smart devices for the purposes of voice-activated dialing and search, as well as on computers and in vehicles. Smartphone voice recognition is well understood, and as already discussed, in-vehicle voice-recognition is another major growth area

As car technology improves, more features will be added to cars and these features will most likely distract a driver. Voice commands for cars allow a driver to issue commands and not be distracted. Most speech recognition software on the market today only have about 50 to 60 voice commands, but Ford Sync has 10,000. Other examples of voice command systems in vehicles include:

- Ford Sync
- Lexus Voice Command
- Chrysler UConnect
- Honda Accord
- GM IntelliLink

A driver may use the feature to look for nearby restaurants, look for gas, driving directions, road conditions, and the location of the nearest hotel. Currently, technology allows a driver to issue voice commands on both a portable GPS like a Garmin and a car manufacturer navigation system.

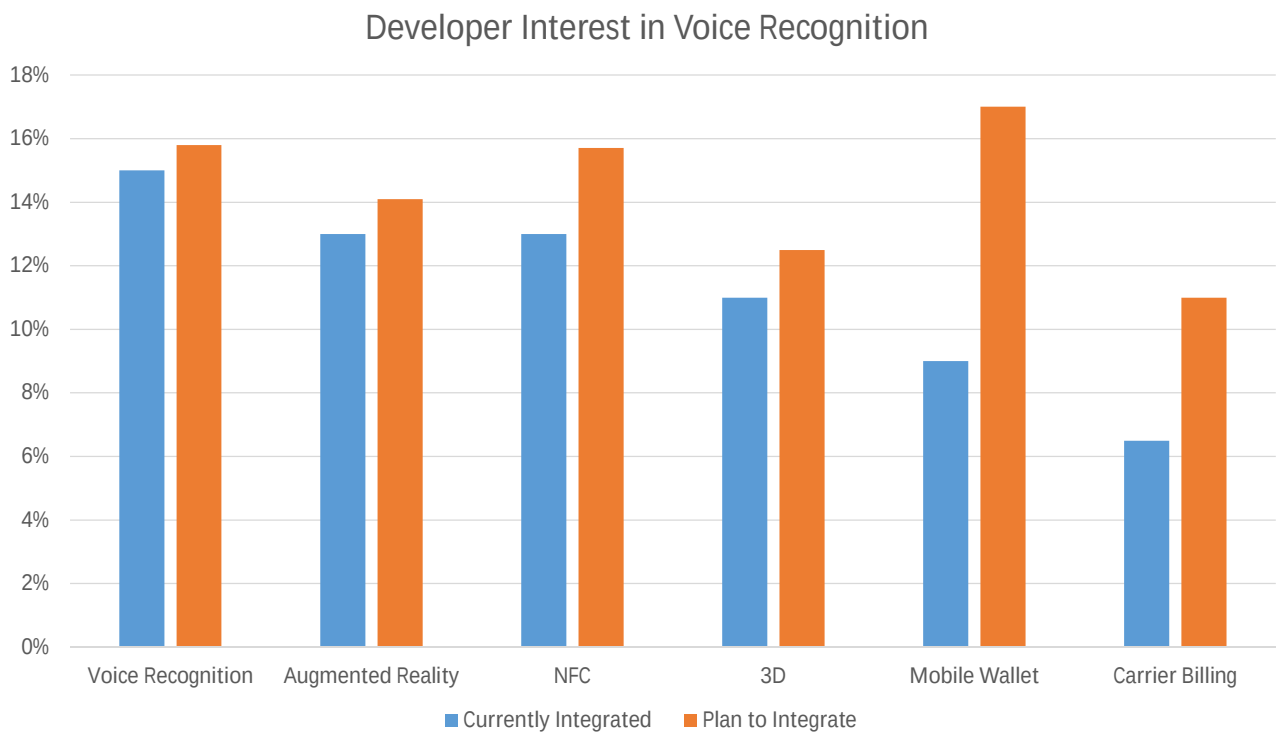
Increasingly basic commands are moving into smart home products and home automation such as LG's home chat service that allows for texting of smart appliances[10] or Dacor's smart oven[11].

Companies have also emerged for developing voice-activated controls to break away from operating system-dominated products offered by major players such as Apple, Google and Microsoft such as Wit.ai[12] (acquired by Facebook in early 2015) to help developers build speech interfaces for their app or device. They offer an API that turns natural language (speech or messages) into actionable data. Developers are using these tools to build home automation products[13]

[10] http://www.techhive.com/article/2023865/ces-smart-appliances-big-tvs-on-deck-at-lg.html
[11] http://www.techhive.com/article/2864387/dacors-android-oven-turns-up-the-heat-with-voice-activation.html
[12] https://wit.ai/
[13] http://hackthenorth.devpost.com/submissions/27003-home-ease

Wearable devices with minimal screen real-estate such as smartwatches that can offer hands-free activation of a multitude of functions, to smart televisions and games consoles represent other obvious use cases for voice recognition functions.

Figure 5: Developer Interest in Voice Recognition



Source: Strategy Analytics

Strategy Analytics research also show that there are multiple other areas where voice integration can simplify and enhance various services, from entertainment to transactional functions using NFC, billing and mobile wallets, for example.

7. Conclusions

As IoT applications proliferate they present new and innovative ways to collaborate, communicate and interact, both from a human and machine perspective. While most of the focus in the IoT has been on data, the integration of voice into IoT applications offers a versatile method to provide human interaction; communication and control, that with proper implementation and consideration for relevance, costs and human factors, can provide a more flexible user experience in a more economical way than traditional methods such as touch screens or data input.

The growth in service providers implementing IMS and the rapid roll out of LTE services in many markets, in addition to existing private VoIP/SIP networks, represents a new and exciting opportunity to cost effectively integrate voice into different IoT applications. Naturally, not all IoT applications will merit voice integration, but there are a multitude of existing and emerging applications that would benefit from voice integration capabilities. And as LTE networks become ubiquitous, the benefits of all-IP connectivity for both data and voice in IoT applications is expected to make VoLTE a popular choice for providing voice services in a wide variety both consumer and industrial IoT applications.

Speech is a singularly efficient way for humans to express ideas and desires. IoT represents a new frontier in integrated voice and speech recognition and telephony applications.

Strategy Analytics believes that voice has a pivotal role to play in the Industrial IoT segments on the shop floor where automation, robotics and mobility will factor in. Specifically, Strategy Analytics projects that voice could capture up to 12% Industrial IoT applications by 2022 and be especially prevalent in the lucrative Asia-Pacific and North American regions.

In the consumer segment, voice will become especially pervasive in wearable IoT and the consumer healthcare segment, Smart Home, consumer electronics and Smart Buildings. In these vertical industries, voice has the potential to capture up to 18% of connections in the 2020 to 2022 timeframe.

Strategy Analytics believes there are four key reasons that voice is suited to a range of IoT applications:

1. **Speech is the natural mode of communication for humans.** It is both intuitive and easier to convey commands verbally.

2. **Voice recognition is particularly appealing when the human's hands or eyes are otherwise occupied**. For example it may not only be convenient but also a legal requirement to use verbal commands (if any) rather than be distracted by typing or using other means of communication, as in distracted driving legislation, for example.

3. **Voice telephony is an efficient means of bi-directional voice communication with machines that can listen, and respond without the need for complex commands.**

4. **Cost saving factors: Voice integration could potentially challenge the need for a touch screen on many devices,** as it reduces the cost for devices that will be dormant for the majority of the time. There is a point at which devices that are not used frequently (e.g. a few times a year), do not need a touchscreen and it would be more cost effective to simply offer connectivity to a call centre to achieve the same function. This represents a change from a volume to a transactional model e.g. a million devices that generate 50,000 calls per year are probably better serviced through a voice connection than putting a $5 dollar screen on a million devices. There are exceptions to this, where the device is not a basic terminal and the users has a smartphone or wearable for a screen.