Independent Investment RESEARCH

SYNC.

Global investment themes: technology, media and telecoms

Issue No. 142

20 December 2016

Voice The next computing platform

Voice is set to become the next big computing platform

Many of us will be talking increasingly to machines. Products such as smartphones, in-car infotainment systems, biometric ID systems, home speakers, games consoles, televisions, doors, fridges and industrial machines are already, or will soon become, "conversational".

Speech recognition technology is currently about 90% accurate. In order for voice-activated products to go mainstream, speech recognition needs to reach 99% accuracy. This may take until 2030.

But, in the meantime, investors need now to factor voice-driven AI into their thinking and into the way they assess the future prospects of many of their mid-term investments.

How do investors play this theme?

For investors, most of the value lies in four critical segments:

- **Semiconductors:** Voice technology requires chips optimised for handling fast streams of large datasets and parallel processing. This should benefit Nvidia and AMD, leaders in GPUs and Xilinx, the leader in programmable chips. Intel's Altera unit should benefit too if Intel manages integration better. KnuEdge is the tech start-up to watch.
- **Artificial intelligence:** Deep learning, a type of machine learning using artificial neural networks (computer programmes built to mimic the complexity of the human brain), is the AI technology that works best for developing the algorithms needed for speech understanding. The current leaders are Google, Baidu, Amazon, Microsoft, IBM and Facebook, with Google in the lead. Vicarius is the start-up to watch.
- **Voice APIs:** Voice application programming interfaces (APIs) allow manufacturers or internet companies to plug their products and services into a ready-made voice platform. Nuance has long been the leader in this space. But, now, the Internet giants are offering "voice APIs" from their Clouds. Nuance is a likely bid target for any technology company that has fallen behind in voice. iFlytek is China's champion. Mindmeld, Sensory, Theatro, VoiceBox and Lingban are on our start-up watch list.
- **Voice-driven products:** Amazon's Echo, a voice-activated intelligent home speaker, leads this space. In time, many "things" will have "conversational technology" built in. Beneficiaries of this trend are likely to be those who provide the AI platforms that power these voice-driven products: Amazon Alexa, Google Assistant, Baidu Duer, Microsoft Cortana, Apple Siri, Facebook M and Samsung's Viv.

Inside

Report type

- **Single theme**
- Multi-theme
- Sector Scorecard
- Best Ideas

Cyrus Mewawalla
Managing Director, Research
cyrus@researchcm.com
+44 (0) 20 3393 3866

Mike Orme
Senior Analyst, Research
michael@researchcm.com
+44 (0) 20 3393 3867

Elgen Strait
Managing Director, Sales
elgen@researchcm.com
+44 (0) 20 3744 0105

www.researchcm.com

Contents

Players

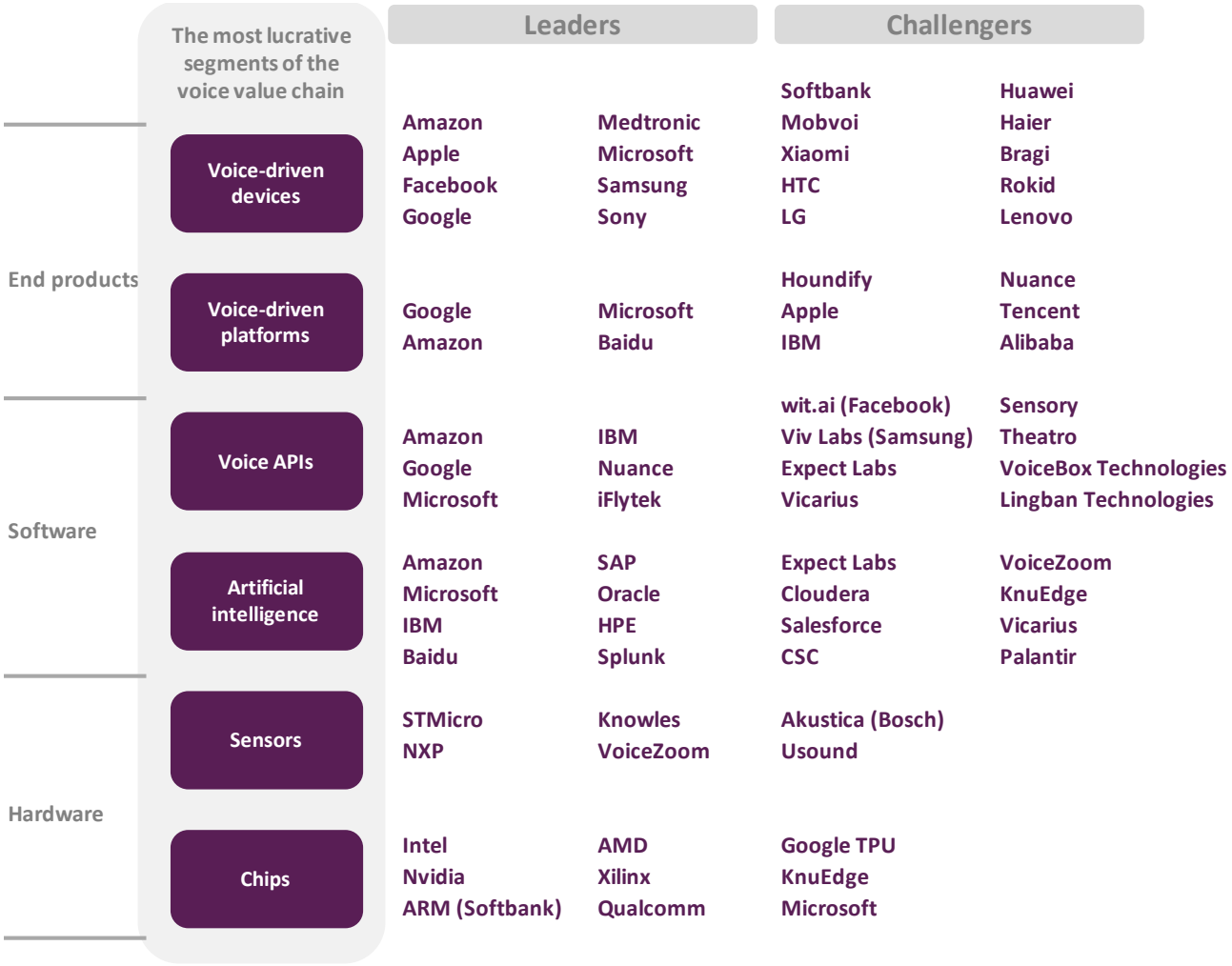
In the movie, *Her* (2013), a guy falls in love with a disembodied voice and confides everything to her. She finally gets so bored with this babbling, pathetic, self-centred human that she leaves him in the lurch to engage with other Cloud-based entities.

Many of us will be talking increasingly to machines. Cars, biometric ID systems, household equipment, games consoles, augmented reality head-wear, industrial machines and robots will all become voice activated within five to ten years' time.

This report looks at the prospects for voice technology and what it means for investors.

The chart below lists the most lucrative segments of the voice value chain, where investors should be positioned. We highlight the leaders and challengers in each of these key segments.

Where should investors be positioned to benefit the most from the "voice" investment theme?
Below, we list the most lucrative segments of the voice value chain



Source: CM Research

Trends

Here are the key themes covered in this report.

Trend	What's happening?
Voice hype	Voice will become one of the key investment themes of 2017. Expect more headlines in the business media covering the importance of "voice" technology as a key determinant of the future. Voice-activated products will emerge in the fields of biometric ID, ADAS (advanced driver assistance systems) in cars, virtual office assistants, retail shopping, gaming, the intelligent home and robotics. The products that are likely to popularise voice the most are Amazon Echo (a home speaker powered by Alexa, a voice-controlled AI service, launched in 2014), Google Pixel (a smartphone designed by Google with a voice controlled virtual assistant, launched in 2016) and Google Home (a copy of Echo, to be launched in 2017).
Accuracy of voice algorithms	Speech recognition technology is currently about 90% accurate. In order for voice-activated products to go mainstream, speech recognition needs to reach 99% accuracy. So, much of the focus of R&D in the AI field will be on getting "conversational tech" to higher levels of understanding and insight. This will require the development of more powerful algorithms, notably learning algorithms based on artificial neural networks, working on richer and more voluminous data streams and concept libraries. This data will be generated as more Internet users switch to "voice" as their preferred method for searching, browsing, issuing commands to "listening" equipment such as Amazon's Echo and interacting with virtual office assistants like Apple's Siri. Baidu, Google, Amazon and Microsoft are the leaders here.
Deep learning	There will be a Cambrian explosion in "deep learning" (a type of AI technology), due to continuing advances in computing power, new chip designs, virtually unlimited availability of online data and the development of complex algorithms enabling computer systems to train themselves. It will bring the 99% accuracy target ever closer in voice recognition and appreciably enhance speech understanding. For now, Google is the clear leader in deep learning. (See "Artificial Intelligence – Vol. II")
Skills shortages	Amazon, which already has over 1,000 people working in AI – substantially in voice-driven AI – has been scouring the market for another 500 people of late, Meanwhile, Apple, which needs to build up its AI resource, has been finding it difficult to recruit the staff it needs. This is putting ever increasing pressures on the global recruitment market for already scarce talent in maths, data science and AI.
Data centres	As and when voice really takes off, future data centres could hit a wall. They are already stressed – most are expensive, take up a lot of real estate and are energy intensive. Voice will add huge volumes of extra data to deal with. This prospect is a major added force in a reshaping of the semiconductor industry away from general-purpose processors to new chip architectures based on less power-hungry graphics processors (GPUs) and programmable chips (known as FPGAs). GPUs offer the speed necessary for the "learning" function of speech recognition, while FPGAs are better for "inference" (because they can run the huge numbers of tiny calculations in parallel that "voice" processing demands). While GPUs – notably from Nvidia – are currently the chips of choice for parallel processing and neural networks, Microsoft is shifting to FPGAs – where Xilinx and Intel are the leaders – and Google has designed its own custom chips called Tensor Processing Units (TPUs). Interestingly, a NASA spin-out called KnuEdge has developed a neurobiological chip for voice-specific neural computing.
Mobile platform voice wars	A core contest is building up between Google Android, Apple iOS and Amazon Alexa, with each striving to gain the widest spread of "voice" apps – especially in the fields of in-car infotainment, the smart home, virtual office assistants and augmented reality.

Trend	What's happening?
AI-first companies	Just as many Internet companies in the last decade designed their service to be "mobile first", so many technology companies are now fixated with being "AI first". Apple, Amazon, Baidu, Google, IBM and Microsoft are among them. Google leads the pack with its concept bank of nearly a billion "concepts" for understanding the world and the sheer scale of its server farms. Baidu, China's Google, is the clear leader in China and has a world class AI operation focused on natural language processing based in Silicon Valley. Like Google, Baidu is also a leader in developing the "brains" for self-drive vehicles with "voice" playing a critical role.
Internet companies	The leaders in voice currently consist of Google, Amazon and Baidu in China with Microsoft expected to start to cut a dash next year, not least because its partially voice controlled HoloLens augmented reality system is due to be rolled out in 2017. Microsoft benefits from its high R&D spend in the area, its Azure Cloud operation plus its huge base of Windows users who provide data for improving Cortana, its voice-based office assistant. Apple, Facebook and Samsung are all trying hard to catch up, with each having acquired a start-up in the field: Samsung bought Viv Labs, Facebook acquired wit.ai and Apple purchased Turi. Facebook, which says it's not actively working on voice, for now, can already claim leadership in the critical area of convolutional neural networking.
Voice APIs	AI is expensive to develop in-house. So, the big Internet companies, seeing an opportunity, have already started to offer Cloud-based AI services for rent. These include Google TensorFlow, Microsoft Azure Machine Learning, Amazon ML, IBM SystemML, GE Predix Cloud and Baidu WARP-CTC. Now these same companies are developing full voice applications accessible from their Clouds too. However, a new breed of AI companies – such as MindMeld, Theatro, Sensory and Vicarius – is developing application programming interfaces (APIs) that allow software developers to integrate voice technology into a variety of applications. Nuance, the incumbent player in this field, is likely to get acquired by one of the big technology giants playing catch-up in voice. These specialist companies offer deeper domain expertise in certain sectors than the "generalist" voice APIs currently offered by Amazon, Baidu, Microsoft, Google and IBM.
Enterprise software	The leading enterprise software suppliers – Oracle, SAP, IBM and HPE – all realise that voice is the next computing platform and are scrambling to get in the game. At a minimum, they need to make their enterprise resource planning (ERP) systems and other Software-as-a-Service (SaaS) services accessible by voice. There is rising speculation that IBM will make marked progress with Watson operating from its Bluemix Cloud to offer voice-driven deep domain expertise in medicine and within the industrial Internet.
China	China is likely to become the first mass market for the deployment of voice technology. For Chinese consumers, voice technology is particularly attractive because of the difficulty of typing Chinese characters into a keyboard – few know Pinyin phonetic Chinese. China's 250 million or so tech-savvy, urban millennials are likely to be early adopters of voice technology. Hundreds of AI start-ups – many of whom are funded by the BATs – are fighting it out to become the "voice" standard adopted by Baidu, Alibaba or Tencent (BAT). Meanwhile, Baidu is the only BAT with world-class AI and "voice" status. Its Deep Speech 2 platform has proved a better Mandarin transcriber than competent humans and its Duer voice assistant for search is catching on. iFlytek, the State-supported voice platform has forged a close alliance with Israel's VocalZoom. Huawei, Lenovo, Xiaomi, among others, may well look to acquisitions to beef up their "voice" resource.
Customer service centres	Many multinational companies have moved their customer service operations – which are labour intensive – to low wage countries to save costs. If voice takes off, vast swathes of people in these customer service centres dotted around the world may be made redundant and their function replaced by chatbots. A chatbot, or "chatter robot", is a computer program that is intelligent enough to have a conversation with a human using voice or text.

Source: CM Research

Industry analysis

"We'll soon be having more conversations with 'bots' than with our spouses."
Daryl Plummer, Gartner Fellow and Chief of Research for Cloud computing

"I hope someday to have grandchildren who are mystified at how, back in 2016, if you were to say 'Hi' to your microwave oven it would rudely just sit there and ignore you."
Andy Ng, Chief Scientist, Baidu

Voice platforms are approaching take off point…

Back in the 1980s, Bill Gates, co-founder of Microsoft, was saying "voice" would be the next big thing in computing.

Nearly 40 years later, it is poised to become so. In the process, voice platforms will transform the way we interact with an ever-swelling array of connected digital devices. The way we work and play in ten years' time will be very different from today.

The first serious exercises in modern, voice-driven artificial intelligence (AI), such as Apple Siri, Microsoft Cortana and Google Now were much hyped, but proved somewhat underwhelming. A new generation of voice platforms led by Amazon Alexa, Google Assistant and Baidu Deep Speech 2 has upped the ante and set a more demanding pace.

Billions of dollars are being invested by the Internet giants and the leading electronics manufacturers such as Samsung and LG to usher in the age of voice-driven AI.

The enabling technologies – chip architecture, big data analytics, machine learning and Cloud services – are finally working in combination to crack the outstanding problems in what is arguably the most complex and difficult area of computing.

Nonetheless, there is still more work to do before "voice" becomes a mainstream computing platform. In particular, there is more heavy lifting in AI and compounding build-ups of ever larger datasets from which algorithms can learn to become smarter. On present progress in machine learning, this will probably be achieved by 2025.

As investors found with smart watches – and many other forms of wearable technology – a new ground-breaking technology can create global hype. Yet, without a strong use case, it can very quickly become a flop.

For voice-driven products to become ubiquitous, not only must they work seamlessly, but users need to be given a reason to buy them.

For now, credible use cases include home speakers (e.g. Amazon Echo), personal robots (e.g. Softbank Pepper) and in-car infotainment systems.

…so investors need to factor voice-driven AI into their thinking

Well before 2030, millions of us will be using "voice" to interact intelligently with a lot more than the Amazon Echo. Intelligent machine applications involving speech recognition and natural language processing will scale up fast to cross the chasm into mass markets.

By then, voice-driven services will become part and parcel of smartphones, smart glasses, home hubs, kitchen equipment, TVs, games consoles, thermostats, in-car systems and apparel. And many "voice-driven" systems will work in conjunction with virtual reality and augmented reality platforms.

A new and evolving technology stack will enable it – ranging from new chip designs to huge data centres, Cloud-based voice application engines, downloadable APIs and convolutional neural networks.

Voice platforms will become a driver of safety and productivity in the workplace: virtual personal assistants will make office workers more efficient; collaborative robots will become more responsive to their human co-workers in factories; maintenance workers will become more productive if they can talk to the machines they are monitoring; and banks will be able to use voice as a type of biometric ID to authenticate access to a customer's bank account (replacing user names and passwords).

Step changes are in process, notably in deep learning, a branch of machine learning where the structure of the algorithms is inspired by the neural networks that make up a human brain (see "Artificial Intelligence – Vol. II" for more detail).

Deep learning programmes enable software engineers to discover patterns in huge volumes of fast moving data. They allow voice-controlled agents to understand ever more about the user and infer intents, meaning, context and objectives. Deep learning platforms offer anticipatory advice and insight as they "listen" and "learn". The more data they are fed, the more they themselves gain in intelligence and knowledge.

It is a case of AI driving more AI.

China's role in voice platforms will be big

China is likely to lead the world in the adoption of voice commanded systems at scale for three reasons. First, the Chinese language – which requires knowledge of around 2,000 different characters for basic literacy – is not that suited for a computer keyboard; second, China's ten year plan – known as "Made in China 2025" – considers next generation technologies such as artificial intelligence a key determinant of China's industrial strategy (see "China Tech"); third, its massive tech-savvy population provides huge datasets with which to power deep learning systems.

China already has hundreds of start-ups in the field of voice-controlled AI platforms, funded by a combination of the BATs (Baidu, Alibaba and Tencent), a burgeoning venture capital industry and the government itself.

Voice algorithms are still only 90% accurate

Voice recognition needs to become 99% accurate for "voice" to become ubiquitous. This compares to 90% accuracy today and under 80% back in 2013. It may not get there. There is much work still to do in deep learning and there may be social blowbacks against people who walk down the street shouting at their headsets.

Nonetheless, voice is an investment theme that investors must start thinking about now because of the evolving business context in which it operates.

The evolving business context around voice

Investors need now to factor voice-driven AI into their thinking and into the way they assess the future prospects of many of their actual and potential mid-term investments. By 2020 at least 50% of the Forbes Global 2000 list will have to make conversational technology a key component of their offerings, according to IDC.

Moreover, these leading companies will need to be adept at using key platforms such as Cloud-based big data analytics, social networks and mobile technology as default choices for their customers.

The innovation accelerators will be AI, the Cloud, the Internet of Things (IoT), augmented and virtual reality, robotics, 3D printing, next generation cybersecurity and blockchain. Many of these platforms will inevitably be voice-controlled.

This will call for new leadership benchmarks and metrics for analysts to assess companies and their growth prospects such as digitisation of offerings, customisation, operational efficiency and agility. Voice, again, will play a pivotal role.

IDC reckons a third of the leading players in every industry will fail in this.

By 2020, however, the boards that run major corporations will be filling with a new generation of CEOs and COOs, many of whom will have spent at least five years in tech leadership roles and who will intuitively understand the importance of voice.

These are all factors for investors to consider when picking out the winners from the losers.

Value chain

Voice is not really an "end product" that consumers or businesses can purchase. It is, however, the computing interface on which many of us are likely to interact with the next generation of advanced technologies. Products such as smartphones, in-car infotainment systems, biometric ID systems, home speakers, games consoles, televisions, industrial machines, augmented reality glasses and personal robots are already, or will soon become, "conversational".

Because voice is not a tangible product or service, the value chain for "voice" is difficult to visualise.

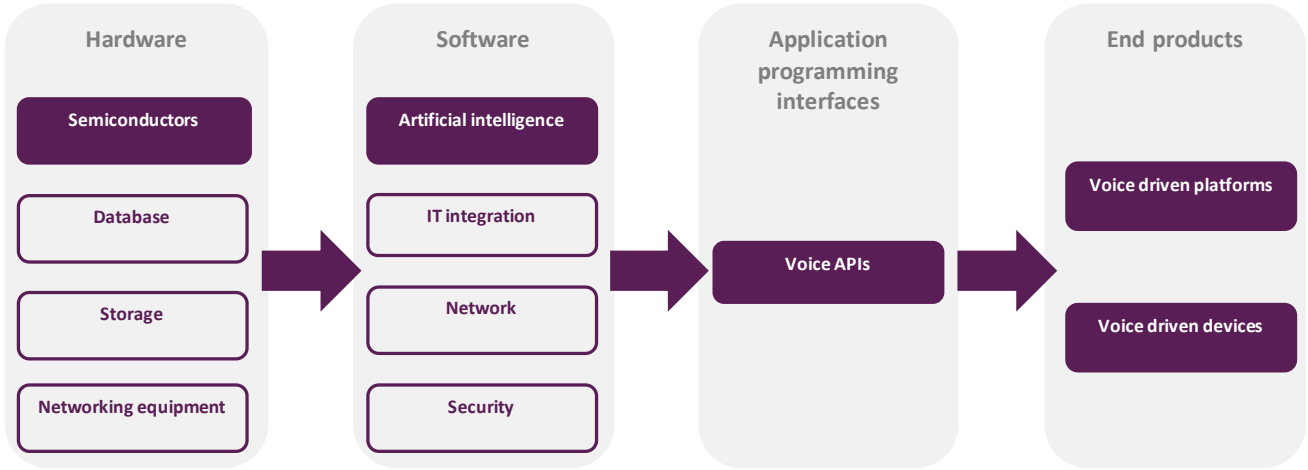
Nonetheless, identifying the key segments of the voice value chain can help investors focus on the most lucrative elements within the *voice* investment theme.

Most of the value lies in four critical segments.

Within hardware, the key differentiating technology for voice is the custom-built chip. Within software, the key differentiator is artificial intelligence. Voice APIs (application programming interfaces) are a critical component of the value chain, because they are the key enabler for products and services to become "conversational". Finally, within end products, the scope for voice platforms to enhance value is almost limitless.

The "Voice" value chain is difficult to visualise…

…but most of the value lies in four critical segments: chips, AI, voice APIs and voice-driven platforms and products.



Source: CM Research

Of the larger technology stocks, Google, Amazon, Microsoft and Baidu have the most exposure to the "voice" investment theme. The two companies closest to pure plays on this theme are iFlytek and Nuance Communications. Thus far, Apple appears to be less advanced in voice than we would have expected.

In the unlisted world, the leaders in voice include Expect Labs/MindMeld, Cloudera, KnuEdge, Mobvoi, Vicarius, VocalZoom and VoiceBox Technologies. Each is in the vanguard of creating the new order needed to make "voice" the next big thing in computing. For Apple, Facebook, Alibaba and Tencent – all of whom have fallen behind in "voice" – these privately held companies would be highly attractive.

Samsung, which is snapping up virtually anything that moves in the field – it recently acquired Viv Labs and Harman – these voice companies would make an interesting bolt-on.

Other technology companies that should benefit from a take-off in "voice" include Xilinx, Nvidia, Micron, Splunk, SAP, and IBM.

Hardware

The key differentiating piece of hardware for voice is the custom-built chip.

Voice technology requires chips optimised for deep learning (a type of machine learning) and the parallel processing of fast non-volatile memory in neural networks that approximate to the workings of the neo-cortex of the human brain.

Conventional serial processors such as Intel microprocessors are not fit for purpose.

More appropriate are graphics processor units (GPUs) from the likes of Nvidia and AMD and Field Programmable Gate Arrays (FPGAs), notably from Xilinx. The former offer the speed necessary for the "learning" function and the latter are better for "inference".

So far, there is no system-on-a-chip (SoC) offering equal strength in both. But NASA spin-out KnuEdge, backed by $100 million of seed money, claims to be close to bringing a step change system strongly based on neurophysiology to market.

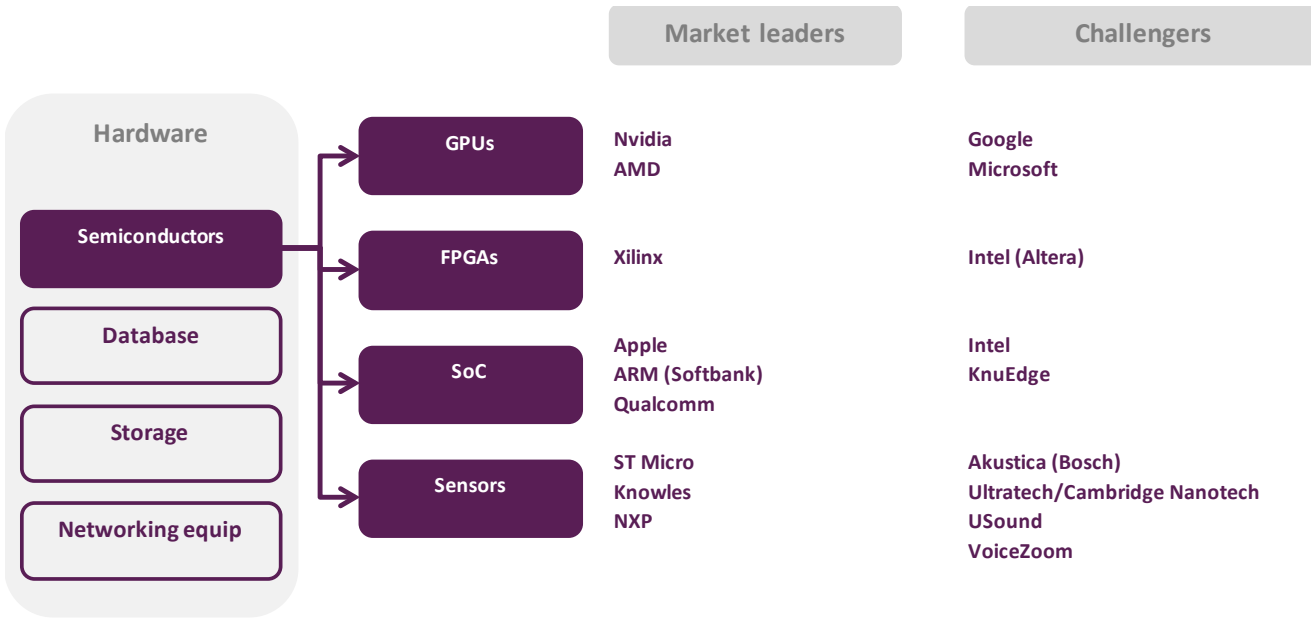
Meanwhile, in 2015, Intel acquired Altera to become a serious contender in the FPGA market, Google has designed its own custom circuit, the Tensor Processor Unit (TPU), to run its open-source TensorFlow machine learning software, and Microsoft has invested heavily in developing its own FPGA capability.

On the memory front, the world is waiting for 3D XPoint – a new type of non-volatile, flash memory developed jointly by Intel and Micron – that claims to be cheaper and faster than anything before it. The commercial launch of 3D XPoint has been delayed until late 2017, at the earliest. Given the vast amounts of real-time data processing and machine learning involved, 3D XPoint will be a game changer for the voice industry.

In terms of voice sensors, ST Micro, Knowles and NXP lead in this field, in general, but are being challenged by VoiceZoom, USound GMBH, Ultratech/Cambridge Nanotech and Akustica.

Voice value chain: hardware

Semiconductors are the critical bit of hardware



Source: CM Research

Software

The key differentiating software for voice is artificial intelligence.

Google's very public pivot to an "AI first" strategy in October 2016 signals a future where the recent rule of thumb for tech start-ups of "Take X and add AI" has been inverted to read "Take AI and build X around it".

Over the next five years, as speech becomes the predominant user interface with connected computing devices – whether robots, cars, home hubs, games headsets, medical equipment or augmented reality glasses – algorithms that recognise images, gestures and speech will provide the keys to unlocking the future.

Thanks to new chip architectures, abundant computing power, high-speed networks, the availability of huge datasets, software analytics is now on the cusp of delivering computers that "understand humans" and engage intelligently with us via "speech".

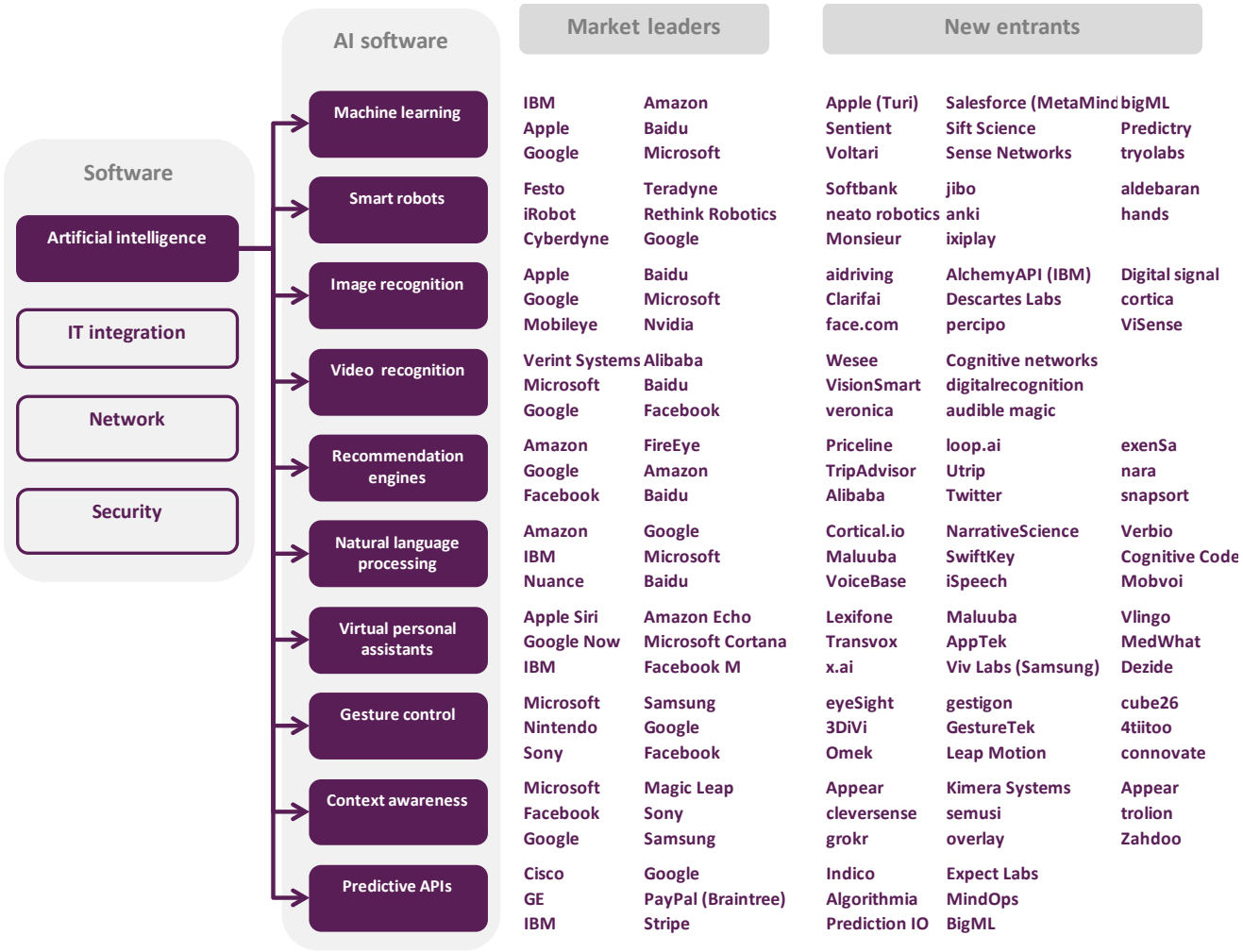
Increasingly, this involves deep reinforcement learning based on neural networks which are virtual models of how the human neocortex operates. The current leaders in this are Google, Baidu, Amazon, Microsoft and Facebook, with Google the pack leader for now.

But an ever-growing pack is catching up as advances in neuroscience are shaken down into software.

Artificial intelligence is key to making voice the dominant computing platform of tomorrow. As we explained in "AI (Vol. II)", there are ten key AI technologies, shown in the diagram below. For voice, the most important ones are machine learning, natural language processing, virtual office assistants, context awareness and predictive APIs.

Voice value chain: software

AI is the critical bit of software



Source: CM Research

Application programming interfaces

Not every company can develop their own voice platforms – they are prohibitively expensive. Voice APIs allow manufacturers or service providers to plug their products and services into a ready-made voice platform.

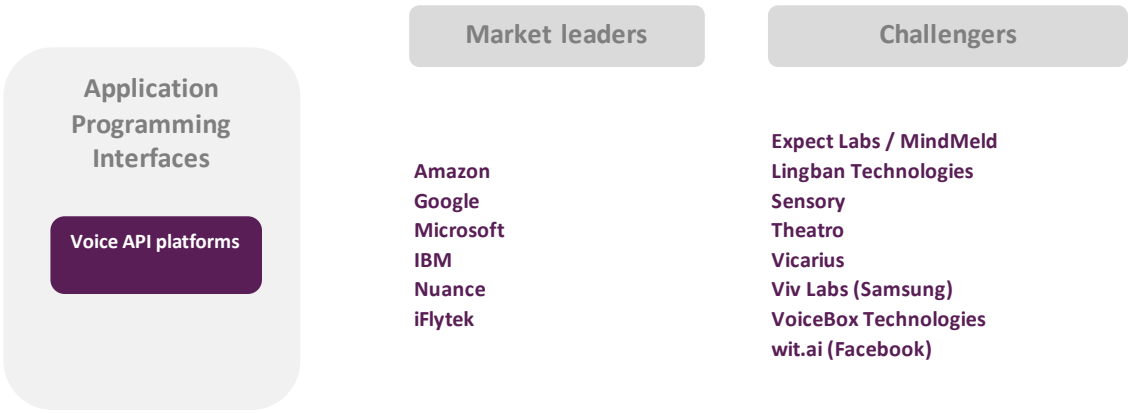
For years, Nuance Communications has been the leading provider of voice platforms as an outsourced service. Now, the Internet giants – all of whom are leaders in speech recognition – are offering "voice APIs" from their Clouds for companies large and small to work with to develop voice capable products and services. These services are incorporated within Google TensorFlow, Amazon Web Services ML, Microsoft Azure ML. IBM offers its System ML.

Whilst these Internet giants offer a generalised voice API, some AI start-ups, such as MindMeld, Theatro and Sensory offer what they claim are "deeper domain" voice APIs (i.e. voice APIs that are custom designed for a specific industry vertical, such as the retail market or the car market).

However, over the next five years, the Internet giants are expected to offer more advanced AI from their Clouds.

Voice value chain: Application programming interfaces (APIs)

Voice APIs enable manufacturers and Internet service providers to make their products and services "conversational".



Source: CM Research

Voice-driven platforms and devices

The scope for voice within end-user devices is almost limitless.

Virtual office assistants in PCs and smartphones are so far the most popular voice-activated services. Market leaders include Apple Siri, Microsoft Cortana, Google Assistant and Baidu Duer.

In the home, Amazon leads with Echo, a home speaker powered by Alexa, Amazon's voice-controlled AI service operated from the Cloud. Google Home is a recently launched rival speaker where "voice" will become the primary interface for the connected home as more household equipment such as thermostats, TVs, fridges and music systems are suitably adapted. In addition, Apple, Samsung and LG among others will be big beneficiaries in the "smart home" market.

Cars, too, are fast adopting intelligent voice-driven systems for infotainment and location based search. Ford's SYNC (which allows drivers to make hands-free calls or control music solely with voice commands) aims to use voice technology to enhance safety and minimise driver distraction

In the field of biometric security, some banks – including Barclays, HSBC and Citi – have introduced voice authentication to some customers to reduce reliance on passwords and secret questions, which can be easily hacked.

Audio and visual entertainment systems are now beginning to use spoken voice for content discovery, on devices such as Apple TV and to issue in-game commands on games consoles from Sony and Microsoft. In augmented reality, a new breed of voice-controlled AR devices will come to market with Microsoft's HoloLens in the vanguard.

In industry, voice-controlled devices will dominate marketplaces that require hands-free mobility: hospitals, warehouses, laboratories and production plants.

We list some of the leading players below.

Voice value chain: end products

Note: the extent to which voice is deployed is hypothetical in some of the references below.

		Market leaders		Challengers	
End products	Virtual assistants	Apple Siri	Google Assistant	Microsoft Cortana	Viv Labs (Samsung)
		Baidu Duer	Amazon Alexa	Facebook M	Alibaba
Voice driven platforms	Car infotainment systems	Nuance Dragon		Harman (Samsung)	
		Blackberry QNX		Houndify	
	Chatbots	Google Allo	Facebook Messenger	Chatfuel	Pypestream
		Apple Siri	Microsoft Tay	JiffyBots	Twyla
		Amazon Alexa	Tencent	Talla	
	Biometric ID	3M		ArmorVox (Auraya)	
		Nuance		NEC	
Voice driven devices	Industrial internet products	IBM Bluemix		SAP	Cisco
		GE Digital		Oracle	Ericsson
	Augmented reality devices	Microsoft HoloLens		Magic Leap	Atheer Labs
		Google Glass 2		Facebook Oculus	Meta
	MedTech devices	Medtronic	Nuance Dragon	Johnson & Johnson	Siemens
		Apple HealthKit	Google	Philips	Olympus
	Home speakers	Amazon Echo		Google Home	
				Rokid	
	Voice controlled games headsets	Microsoft Xbox One S		HTC Vive	Google DayDream
		Sony PS4		Samsung Gear VR	Facebook Oculus

Source: CM Research

Investment timeline

The "voice investment theme" is probably a fifty year cycle that is heavily interconnected with the "Artificial Intelligence investment theme".

A voice technology timeline

Sound became recordable in analogue form back in 1877. But speech recognition did not really begin until the 1990s, when Markov Models were used.

In 2011, Apple introduced Siri as a voice-activated office assistant and in 2014 Amazon introduced Echo, a voice activated home speaker.

In 2016, Google introduced the Pixel smartphone, with Google Assistant – a voice-activated virtual office assistant – at its core.

In 2017, Google Home, a copy of Amazon Echo, is expected to sell well. Also in 2017 Microsoft, will launch the HoloLens augmented reality headset. It will untethered (i.e. it works without being connected to a PC) and it will be partially voice-activated.

By 2020, some voice platforms will reach 99% accuracy, the level necessary for mass adoption.

The story of voice so far

1877	Edison invents phonograph, first device to record and reproduce sound.
1879	Edison invents first dictation machine.
1939	Bell Labs demonstrates first electronic voice synthesiser.
1952	Bell Labs introduces, Audry, machine capable of understanding spoken digits.
1962	IBM unveils Shoebox which understands 16 spoken words in English.
1976	DARPA (US military research agency) develops machine capable of understanding 1,000 words.
1990	Hidden Markov Model begins to be used in speech recognition to predict sounds as words.
1990	Dragon launches first consumer speech recognition product, Dragon Dictate.
1996	IBM launches MedSpeak, first commercial product capable of recognising continuous speech.
2006	NSA uses speech recognition to isolate key words in recorded conversations.
2007	Windows Vista becomes first Windows operating system to incorporate speech recognition.
2011	Apple introduces Siri, its voice-activated office assistant
2014	Microsoft unveils Cortana, its Siri equivalent.
2014	Amazon lauches Echo, a voice-controlled speaker liked to Alexa, its cloud-based AI platform.
2016	Google announces Pixel phone and Google Home - both powered by Google Assistant.
2017	Microsoft brings its game changer, HoloLens AR, to market with Magic Leap nowhere to be seen.
2020	Some voice platforms achieve 99% accuracy.
2021	People start shouting commands into their apparel and jewellery as they walk down the street.
2022	A raft of new neurophysiological-based systems are deployed to yield step changes in speech recognition and understanding.
2025	"Conversational technology" becomes ubiquitous.
2030	IT systems incorporating biological subsystems make computers even more human like.
2030	Quantum computers change everything.
2050	50% probablity that computers will surpass human intelligence levels

Source: CM Research

To help investors visualise where we are in the investment cycle, let's look at some statistics.

Voice by the numbers

Here is a selection of statistics setting out the current state of play of voice technology:

- Today, 10% of Internet searches worldwide are initiated by a voice query (Google).
- By 2020, over 50% of Internet search queries in China will be voice-controlled (Baidu).
- Within five years, half of all global web searches will be by voice or image (Mary Meeker of Kleiner, Perkins, Caufield and Byers).
- By 2020, 30% of all web browsing will be done by voice (Gartner).
- 65% of US smartphone owners use voice-controlled virtual office assistants, such as Apple's Siri.
- Apple's Siri gets over a billion queries per week.
- Over a quarter of Microsoft Windows 10 taskbar searches are voice-driven.
- Within 10 years, people will have more conversations with "bots" than with their spouses (Daryl Plummer, VP at Gartner)
- Within 7 years, voice recognition will achieve 99% accuracy – up from 90% today and 80% in 2013.
- Amazon has over 1,000 people working on "voice" technology and may add a further 500.
- Within 10 years, "conversational technology" will become ubiquitous (Mary Meeker of Kleiner, Perkins, Caufield and Byers).
- Much of the wearable technology market – worth $34 billion by 2021 – will include voice-driven devices (CCS Insight).

Technology briefing

Voice recognition and speech understanding present a significant challenge to software engineers, particularly because they are interdisciplinary.

The world's best computer scientists, mathematicians, linguists and neuroscientists have been thrown by the difficulty in deciphering different types of "phones" (a linguistic term which refers to distinct speech sounds or gestures).

Listening is much harder than it looks (or sounds) because of the following factors:

- background noise;
- the speed with which some people speak;
- the variety of accents used in each language;
- the tone of the speaker (which can convey the context of the conversation);
- the double meanings of some words (e.g. date can mean a fruit or a rendezvous);
- the different spellings of identical sounding words (e.g. its and it's)

On top of these come the issues of syntax and semantics. The former refers to the sequence in which words are put together to form a sentence (e.g. in English, the usual sequence is subject, verb, and object). The latter refers to the meaning of a particular phrase (e.g. phrases such as "he was in the doghouse" can be confusing to a computer).

Humans find the idiosyncrasies of language relatively easy to pick up – a testament to the power of the human brain – but computers have to be painstakingly trained to learn language as humans speak and understand it.

So getting a computer to pull off anything approaching the human brain's abilities in voice recognition and speech understanding is a non-trivial task.

All voice recognition involves elements of pattern matching, pattern and feature analysis and statistical analysis. Machine learning techniques are often used to speed up the process of "teaching" language to a computer and, increasingly, artificial neural networks are the type of machine learning structures being deployed.

Hidden Markov Models dominate speech recognition…

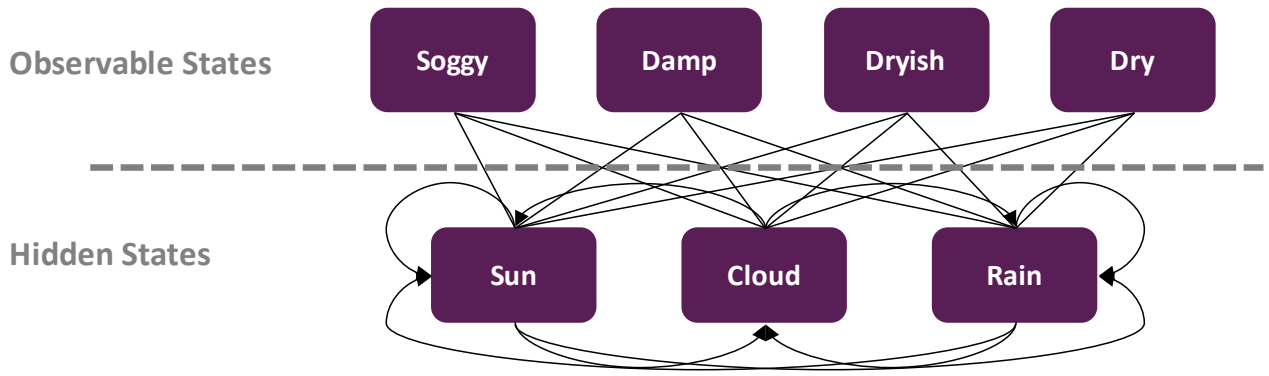
Virtually all modern systems use complex statistical analysis to help figure out what is being said. These statistical models look at the speech data and calculate the probabilities of different interpretations. Several permutations and combinations would be considered, with factors such as which "phone" follows another or silences between "phones" all factored in.

Ultimately, the system builds what is misleadingly called a "hidden Markov model" of each speech segment.

Hidden Markov Models form the basis of computational sequence analysis.

Markov chains are mathematical systems describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event. A Hidden Markov Model is a Markov chain with unobserved (or "hidden") states.

Example: The "states" within a Hidden Markov Model are used to predict the weather



Source: MIT, CM Research

A Hidden Markov Model (HMM) is a statistical Markov model in which the system being modelled is assumed to be a Markov chain with unobserved (or "hidden") states. Markov chains, named after Andrey Markov, are mathematical systems describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event. Thus, Markov chains are constantly hopping from one "state" (a situation or set of values) to another.

HMMs provide a conceptual toolkit for building complex sequencing models simply by drawing an intuitive picture. In addition to being used in speech recognition, they are at the core of a several advanced technology platforms where the rapid interpretation of vast datasets is a critical component, such as genomics, artificial intelligence or autonomous driving.

When it comes to speech recognition, the HMM of each speech segment guesses at which beads are sitting on the string on the basis of what it has gleaned from the sound spectrum and all the "phones" and silences it can reasonably be expected to contain – a so called "Markov chain". This essentially turns speech recognition into mathematically determined "best guesses", confirmed or otherwise by the speaker.

…while Connectionist models created the concept of artificial neural networks

Markov Models have dominated speech recognition since the 1970s, because they work so well.

By contrast, "connectionist" computer models, which came on the scene in the 1980s, created the concept of artificial neural networks, which are computer programmes that mimic how the brain learns to recognise patterns.

Neural networks are vastly simplified digital versions of the brain. They have inputs (for information feeds), outputs (for results) and hidden units connecting the two. If trained with enough "examples", these artificial neural networks learn by gradually adjusting the strength of the connections between the different layers of units.

There is a lot of work being done by scientists to use neural networks side by side with Markov models.

Now the focus is on deep learning

Deep learning is a field of machine learning – a technology which allows AI programmes to learn for themselves. Deep learning systems are built using "artificial neural networks" which model the way neurons in the human brain talk to each other. According to Google, "an artificial neural network is trained by showing it millions of training examples and gradually adjusting the network parameters until it gives the classifications we want." A properly trained neural network can therefore distinguish between signals and general noise.

Deep learning – based on artificial neural networks – is likely to become the AI technology that allows AI systems to reach and then surpass the level of intelligence of the human brain.

With the availability now of enormous amounts of raw computing power to run complex, scientific algorithms fed by huge volumes of fast moving data, processed by parallel processing hardware, the era of Deep Learning systems – key to the 99% accuracy rates required for voice recognition to become a mass market product – has begun.

Deep Learning systems represent a step change from Machine Learning systems, which enable computers to get better, in that they enable computers to train themselves and use multi-layered neural networks and vast amounts of data to do so.

Over the next five years what has been called the "Cambrian Explosion" in Deep Learning will transform voice recognition and speech understanding. A world of "AI First" machines and devices and services will emerge.

What the future holds for voice

As more voice usage data becomes available, speech recognition accuracy and understanding gets better.

The more people use voice interfaces, the more data is gathered and better the algorithms can work, thus delivering greater accuracy, especially with the advent of deep learning algorithms.

Ominously, more and more machines and devices around us will be constantly listening to us to capture data to feed their algorithms. Amazon Echo, the "always on" speaker in your home is an example.

Stunning though the advances from feeding huge amounts of data into neural networks have been, research is underway to make natural language part of a more human-like general and generative intelligence. This uses a categorically different approach: building human-like, virtual brains to become the "CPUs" of AI applications.

Vicarius is a case in point. Backed by a galaxy of technology titans including Peter Thiel, Elon Musk, Mark Zuckerberg and Jeff Bezos, with robot maker ABB as a corporate investor, Vicarious is at the forefront of this new technology.

Vicarious is close to demonstrating a new kind of neural network algorithm to take account of more of the features that appear in neurobiology – such as the ability to picture what the information it has ingested would look like in different scenarios. This enables Vicarius's AI system to use less data inputs and to recognise stimuli and concepts more easily.

Human knowledge, while vast, is not infinite. It is reckoned that between 100 and 500 billion "concepts" or "entities" are useful in human knowledge and Google says it has nearly a billion "concepts" to hand as a start.

Companies section

Listed companies

Our stock watch list below names the companies at the forefront of voice technology. Whilst the "voice" investment theme will have a positive earnings impact on these companies, other investment themes may have a negative impact. So, the stock ratings shown below reflect our assessment of the *net impact* of all relevant investment themes, taking account of valuation and risk. (See Appendix for our methodology.)

Company	Rating	Country	Competitive position in the "voice" industry
Alphabet (parent company of Google)	Buy	USA	At its big event on 4 October 2016, Google made a dramatic pivot to an "AI first" strategy. Everything Google does will now have AI at its core. Its latest products – the Pixel smartphone and the Google Home speaker – will be powered by voice-activated Google Assistant. Given Google's significant lead in distributed computing, this should help move Google to the top of the league in "voice". Further advances in deep learning technology are required before voice technology becomes pervasive in a broad range of digital devices with intelligent conversational powers. Google is the clear leader in the field, although Microsoft has the potential to close the gap fast if it puts enough effort behind it. Meanwhile, Google DeepMind's AlphaGo programme showed earlier this year just how much progress Google has made in deep learning. It has the combined power of huge and fast moving data sets (from Internet search, YouTube, Gmail and Google Maps), parallel processing-memory intensive computing systems at scale and a concept bank of up to a billion of the "concepts" and "entities" that lie behind human understanding of the world. These assets mean that Google is the most likely tech titan to dominate the emerging era of intelligent, conversational technology – whether applied in the home, the car, the office or the factory. But it needs to catch-up with Amazon and Microsoft in the Cloud.
Amazon	Buy	USA	Jeff Bezos, Amazon's CEO, says he has thousands of people working on AI, most notably in "voice" and that he is in the market to hire at least another 500 suitable people. It took over 1,000 people 4 years to develop Amazon's Echo/Alexa product, which continues to impress as a listening and learning technology in the home. It is in the process, with Ford, of moving it into the car. Amazon clearly aims to control the home via Alexa, "speaking" to voice-enabled home devices ranging from kettles to TVs and microwave ovens. Many of the companies that develop such devices – whether Samsung, LG, Haier or Brita – will use APIs, toolkits and full applications downloaded from Amazon Web Services – still by far the world's leading Cloud operation. Samsung and LG are direct competitors in the automated home market and Samsung has just acquired Siri spin-out, Viv Labs to beef up its voice resource. But we believe Amazon is better placed than either of them in voice.
Baidu	Buy	China	China's leading Internet search company is repositioning itself as an "AI first" company. Under the leadership of ex-Google Brain superstar, Andy Ng, Baidu's AI research team – which is centred in Silicon Valley – is developing Deep Speech 2. Like Google's voice platform, Deep Speech 2 uses artificial neural networks. At Baidu, such networks are being fed millions of examples of transcribed speech with the result that the system surpasses humans in accurately transcribing Mandarin. By cracking the appreciable challenges of weather, pollution and noise in China's big cities, Baidu is making fast progress in voice technologies. At the same time, the demand among tech-savvy, mobile-obsessed Chinese urban consumers for voice control is overwhelming, because not many Chinese know (or like) the phonetic Pinyin needed to type into a keyboard. Baidu dominates the protected domestic "Internet search" market (Google is banned from mainland China). By 2020, it reckons 50% of search in China will be directed by voice. Meanwhile, voice-driven AI will be a critical factor in Baidu's well-advertised ambition, like Google, to be a leading contender in the market for self-driving cars.

Company	Rating	Country	Competitive position in the "voice" industry
iFlytek	Not rated	China	iFlytek is the powerhouse in Chinese speech recognition technology. Its competitive position is helped by its status as a National Key Software Enterprise. Moreover, the Ministry of Information Industry of China has appointed iFlytek as Leader of the work group to set up the Chinese speech technology standard. It is also the initiator of China's Super Brain AI project. Unlike Baidu, it markets voice-based machine interfaces to government and private institutions. According to local sources, iFlytek is noted for its natural language processing technology for the educational sector. It has recently formed an alliance with Israel's specialist voice optical sensor company VoiceZoom. It is very much a part of the Communist Party State, having had high profile visits from Politburo chiefs from President Xi Jinping down.
Microsoft	Buy	USA	Microsoft is still benefiting from its position as a momentum story as it focuses on the Cloud, mobility and AI as new layers on top of Windows. Like Google, it, too, has huge sources of data to feed its complex voice APIs and algorithms from its Windows, Office, Xbox and Bing operations to up the ante in voice-driven systems. Meanwhile, its Cortana virtual assistant, is still not impressive when pitted against Amazon's Alexa. Little mentioned, though, is the proficiency of the Skype translation facility. The world is waiting most of all for its augmented reality HoloLens system to move into the mainstream, starting next year and to what extent and how effectively it incorporates voice technology. HoloLens will probably be aimed mainly at Microsoft's billion strong, business user base. In gaming, Microsoft has interesting plans to use voice technology in an upcoming version of Minecraft. Like Google, Amazon, Apple and IBM, Microsoft will be driving to make a big business from providing voice recognition and speech understanding toolkits and applications for developers via its Azure Cloud platform.
Nuance	Buy	USA	Nuance Communications is a pure play voice technology company. It has taken investors on a downward tending roller-coaster ride of late. It is facing rising competition from start-ups like Expect Labs' MindMeld (see the next section covering privately held companies) and from the Cloud giants as they beef up their voice-based offerings. Meanwhile, its Dragon dictation offering for physicians, which enables them to use voice input to capture and document patient-care items in real time on various devices has won high praise. Nuance recently won a contract from TalkTalk to improve the telecom operator's automated telephone system for customer queries. Nuance is a takeover target for some of the larger technology players who have fallen behind in "voice".
Samsung	Sell	South Korea	Samsung is snapping up virtually anything that moves in the field. It recently acquired Viv Labs (voice-based AI) and Harman (big in in-car audio and control systems). The problem with Samsung's acquisition of Viv Labs, however, is that it does not fit with its contract with Google on Android devices. After Samsung's battery-related recall of its Galaxy Note 7 smartphone, its credibility is riding on the commercial success (or not) of the forthcoming Galaxy S8. However, Samsung's contract with Google – which allows it to use the Android mobile operating system – forbids it to compete with "Google Mobile Services" applications such as Google Assistant (a rival to Viv). This means that Viv may never appear on a Samsung-made Android phone. Samsung, of course, has another operating system it has been developing called Tizen, but this system has very few users so building an ecosystem will be next to impossible and, without an ecosystem Tizen phones are unlikely to sell. This goes down the age-old problem with Samsung – namely that it is a hardware company caught in a software revolution. It needs to develop its software ecosystem, but must start from scratch in order to do so.

Source: Company Data, CM Research

Privately held companies

The table below shows some of the strong "voice" players

Company	Country	Competitive position in the "voice" industry
KnuEdge	USA	KnuEdge aims to change how the world addresses large-scale machine learning, signal processing and data analytics workloads. It uses "neurobiologically based" technology. Its custom silicon designs, which have seen first silicon, reflect the way neurons in a human brain become more sensitive to each other as they work more constructively with each other. Under the KnuEdge design, the neurons are connected by connective tissue of an associative routing technology meaning the data can be ferried at warp speed between cores, processors, boards and servers. KnuEdge claims that what it is developing will improve AI response times by faster processors and faster APIs. Meanwhile, KnuEdge has developed neurological-based algorithms for military-grade voice recognition and authentication technology and is selling the technology on the open market. The company, which remains largely in stealth mode, will need to raise more funding. It is the brain child of NASA veteran Daniel Goldin, who was the technology chief on the International Space Station, the Mars robotic explorer and the initiator of the Astrobiology Institute. He is regarded as highly innovative while offering a safe pair of hands.
MindMeld	USA	MindMeld is a speech recognition platform for software developers which has been developed by Expect Labs. Backed by Google, Intel, Samsung, Telefonica, IDG Venture and Greenock among others, MindMeld is widely considered to be setting the pace in enabling companies, large and small, in virtually any sector to deploy conversational AI in their products and services. The company makes clear that its platform based service is not a perfect product yet. What it offers its customers are applications which are improved by user input and interaction. A key feature is that MindMeld's system can interoperate with other voice-based platforms such as Alexa, Siri and Cortana. Its Deep-Domain Conversational AI tool, launched in November 2016, demonstrates deep knowledge and expertise around any custom content domain. This makes it superior in some respects to the voice assistants developed by the big Internet giants. Its clients include the retailer Uniqlo, which is using MindMeld's new AI platform to build highly advanced voice and chat assistants.
Mobvoi	China	Mobvoi is a Google spin-out, based in China. It is developing Chinese language speech recognition for search on wearable devices and makes a voice-driven operating system for wearables. It also makes Ticwatch, a smart watch which competes with Apple Watch in China. Mobvoi is working with Google – currently banned from China – on developing a proxy Google app store for Android wearables in China. Future voice-activated products developed by Mobvoi are likely to involve making payments, authenticating a user by the sound of his voice (e.g. for unlocking doors) and tracking health. It claims to have innovated faster than Android Wear by taking fast feedback from its early adopters and constantly iterating its software at least on a weekly basis. It is also looking at in-car experiences and family robots. Mobvoi has made it clear that it is not a Trojan Horse to be used by Google to get back into China!
Theatro	USA	Theatro is a first voice-controlled wearable device for use by retailers, hotels, restaurants and manufacturing companies. Its software suite of productivity and communication applications accessed through the Cloud are designed to optimize employee, sales, and operational performance.

Company	Country	Competitive position in the "voice" industry
Vicarius	USA	Vicarius is an AI start-up that is building a unified algorithmic architecture to achieve human-level intelligence in vision, language and motor control. Backed by a galaxy of technology titans including Peter Thiel, Elon Musk, Mark Zuckerberg and Jeff Bezos, with robot maker ABB as a corporate investor, Vicarious is close to demonstrating a new kind of neural network algorithm to take account of more of the features that appear in neurobiology, such as the ability to picture what the information it has digested would look like in different scenarios: to learn from less data and recognise stimuli and concepts more easily. Using inductive biases drawn from neuroscience, Vicarius's system requires less data to train a computer than traditional machine learning techniques.
VocalZoom	Israel	VocalZoom is a specialist voice optical sensor company. It has developed a low power optical sensor that measures facial vibrations during speech and combines the added data with output from acoustic microphones. Its customers include Intel, iFlytek, 3M and Microsoft.
VoiceBox	USA	VoiceBox has developed world leading contextual voice technology. It is developing a voice AI platform incorporating its proprietary context management. The company reckons that its competitors pay too little attention to the context of utterances through a conversation. It counts Toyota, Samsung, Renault, Johnson and Johnson and Nuance among its partners.

Source: Company data, CM Research

Appendix: Our "thematic" research methodology

Fundamental equity research does a poor job of valuing technology stocks

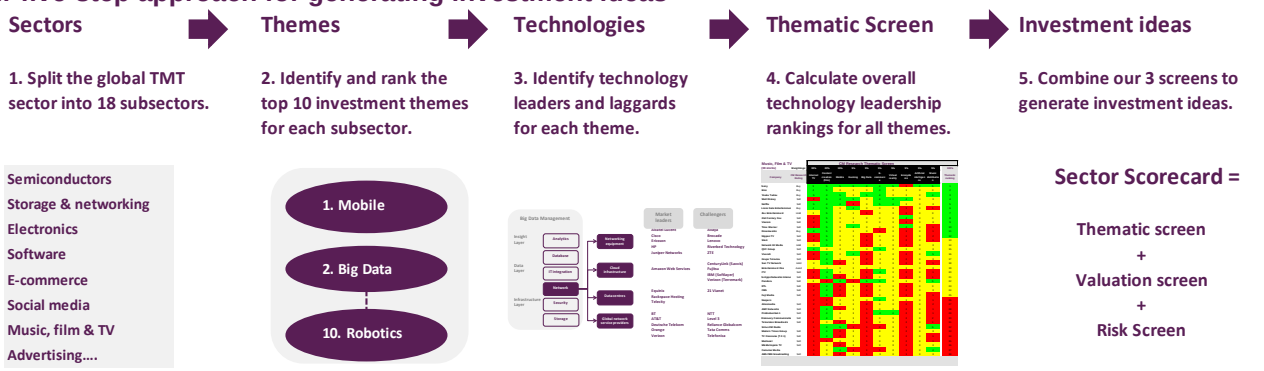
Traditional bottom-up valuation methodologies have a poor track record of predicting share prices in the technology sector. In part, this is because technology cycles these days move pretty fast and it is difficult to judge where you are on the growth curve. Also, valuations tend to be permanently high.

Introducing CM Research's "thematic" valuation approach

So we, at CM Research, have developed an entirely new three-screen valuation methodology for the technology, media and telecom sectors based on a thematic investment approach. This is how it works.

First, we split the global TMT industry into 18 subsectors – ranging from PCs to social media. Second, we identify and rank the top ten investment themes for each subsector. Third, we publish in-depth research on specific investment themes, identifying the technology winners and losers. The problem is that companies are exposed to multiple investment themes, all acting concurrently: some will send a stock up; others will send it down. So our fourth step is to create a thematic screen for each sector to calculate overall technology leadership rankings after taking account of all themes impacting that sector. Finally, we combine this thematic screen with valuation screen and a risk screen to generate a sector scorecard used to create investment ideas.

Our five-step approach for generating investment ideas



Source: CM Research

Each sector scorecard has three screens:

- **The thematic screen** tells us who are the overall technology leaders in the ten technology cycles that matter most to this industry. Each company is scored on the basis of whether we expect earnings in the next 12 months to outperform or underperform consensus numbers on the back of each theme.
- **The valuation screen** tells us which players are the most attractively priced, relative to their peers, using the consensus-based valuation metrics which we believe are most appropriate for each industry.
- **The risk screen** tells us who the riskiest players in each industry are, based on four categories – corporate governance risk, accounting risk, technology risk and political risk.

How our research reports fit into our overall research methodology

We produce four tiers of thematic reports to help our clients select stocks:

- **Single Theme:** These reports offer in-depth research into a specific theme. They identify winners and losers based on technology leadership, market position and other factors. Recent themes include the Internet of Things, Internet TV, Big Data, Artificial Intelligence, Robotics and the Cloud.
- **All Themes:** These reports cover all stocks and all themes within a sector, giving readers a strong sense of how everything fits together and how conflicting themes might interact with one another.
- **Sector Scorecard:** Each sector scorecard has a thematic screen, a risk screen and a valuation screen. The thematic screen identifies overall winners and losers in a sector based on all themes impacting that specific sector. Live scorecards for each of our 18 sectors are available on our client portal, together with our up-to-date stock ratings.
- **Best Ideas Report:** These reports include our high-conviction stock ideas.

About CM Research

CM Research is an independent research provider with a blue-chip list of institutional clients. We analyse emerging trends in the technology, media and telecoms sectors and develop them into global investment themes. We research these themes in detail and then feed the results into a scorecard system to quantify the impact of conflicting themes on individual stocks. Our focus is on disruptive technologies. Our stock coverage includes the top 500 global TMT stocks. Our clients include institutional investors, corporations, consultancies and governments. At a time when many of our competitors have had their reputations mired by conflicts of interest, we fiercely guard our independence. Our service is available exclusively to our clients.

Contact CM Research

Research

Cyrus Mewawalla
Managing Director, Research
cyrus@researchcm.com
+44 20 3393 3866

Michael Orme
Senior analyst, Research
michael@researchcm.com
+44 20 3393 3867

Sales

Elgen Strait
Managing Director, Global Sales
elgen@researchcm.com
+44 20 3744 0105

Matt Barker
Director, US Sales
matt@researchcm.com
+1 646 831 4649

Brian Kern
Director, US Sales
brian@researchcm.com
+1 804 901 2202

17 Savile Row | Mayfair | London W1S 3PN | United Kingdom

www.researchcm.com

Analysts' Certification

The analysts involved in the production of this document hereby certify that the views expressed in this document accurately reflect their personal views about the securities mentioned herein. The analysts point out that they may buy, sell or already have taken positions in the securities, and related financial instruments, mentioned in this document.