



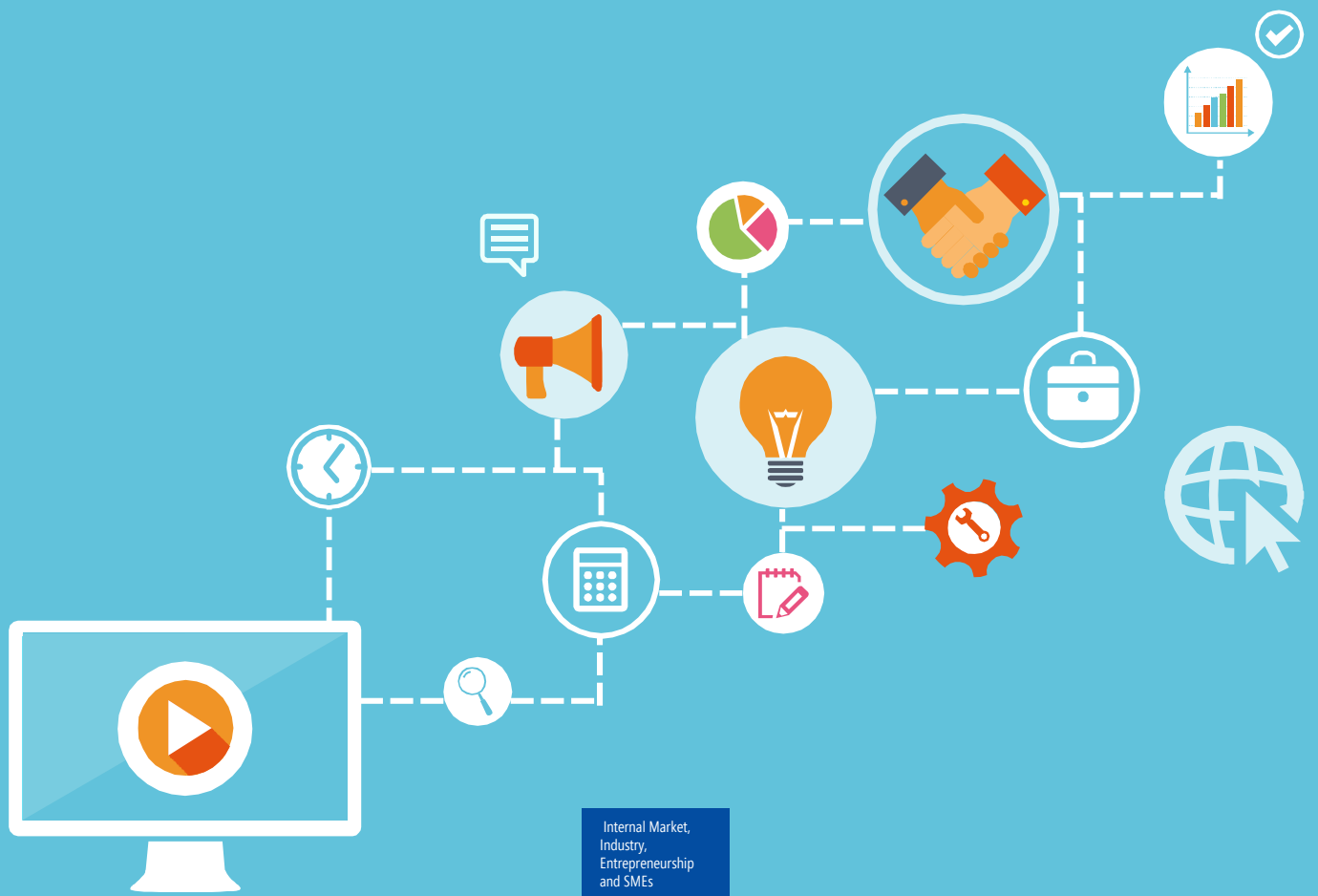
“Hey Siri, wake me up at 7 AM tomorrow”

Talking to Siri is an easier, faster way to get things done. It's always with you — on your iPhone, iPad, Mac, Apple Watch, and Apple TV — ready to help throughout your day. Ask Siri to set an alarm or a destination. Book a ride or a table. Send a payment or a love note. Even change the lighting in your room. And the more you use Siri, the better it knows what you need at any moment. Just say it, and Siri does it.

Digital Transformation Monitor

The rise of Virtual Personal Assistants

January 2018





The rise of Virtual Personal Assistants

The improvements in AI technologies have led to the development of new products, notably VPA, which can now display relatively high accuracy. Internet giants are leading the pack, leveraging their technological expertise, with direct R&D investments, start-ups acquisitions, and access to their own processing infrastructure. These players have notably popularized VPA among their customer base and their devices (smartphones or dedicating speaker) with no clear competitor in sight.

1

VPA usage and technology

Virtual Personal Assistants (VPA) are software program meant to interact with an end user in a natural way, to answer questions, follow a conversation and accomplish different tasks.

Two kinds of inputs are usually possible for a VPA: a voice interface (such as Apple Siri) or a text interface (Google Assistant). The key point is that the end user is supposed to be able to talk to the VPA using his natural language, that is, as he would do to another human being rather than having to use specific sets of commands or a computer language.

Since the launch of Siri by Apple in 2011, the offers of virtual personal assistants have developed rapidly to provide a more generic user interface. This user interface can be accessed from the user device (smartphone or specific device) to perform actions, control objects, answer question and even make recommendations on its own.

Usage of VPA services and devices

Among users, the popularity of VPA is overwhelmingly linked to its hands-free capabilities.

Indeed, more than 60% state¹ that the “usefulness when hands/vision occupied” if one the main reasons for using their voice. In addition, some users’ responses indicate that using their voice seem to provide a better experience overall:

- 30% indicate that voice is a quicker method to search for something
- 24% have difficulties typing on certain devices
- 12% favour VPA to avoid “confusing menus”.

These figures seem to indicate that VPA usage should not be marginalised as a “fad”, since users’ responses indicate that speech recognition systems:

- Are useful in cases where traditional input methods are not available
- Provide, for some users, an improved experience, even when these traditional input methods are available.

Artificial Intelligence is as the basis

The recent development of VPA can be directly linked to recent progress in AI technologies.

Artificial Intelligence can be defined in several ways. Grossly speaking it is the ability for a machine to mimic the cognitive behaviour of a human being,

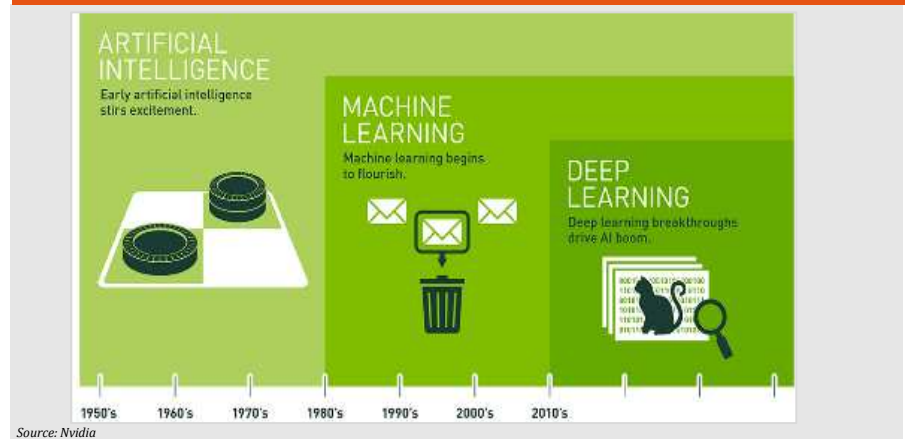
This can however mean several things and can thus be decomposed in a subset of characteristics:

Autonomy: The ability for a machine to act on its own, perceiving its environment, and taking actions to maximize its chances of success in a defined task. The development of these capacities has given rise to the field of automation and applications such as robotics or the internet of things.

Problem solving: The ability for a machine, given defined inputs to produce outputs that maximize predefined criteria of success. The development of these capacities is usually considered to be the field of algorithmic.

Machine Learning: The ability to get better over time at a defined task. Or said differently, the ability for a machine to progress autonomously at problem solving. This is the field that is usually considered as the main topic of Artificial Intelligence nowadays, and indeed most of the excitement around A.I. is focused on Machine Learning.

Figure 1: From Artificial Intelligence to Deep Learning



2

The recent progress in Machine Learning that have made possible the rise of Virtual Personal Assistant, come from a specific approach to Machine Learning: Neural Networks and more specifically Deep Learning Neural Networks.

Indeed, these new approach benefit the many different tasks which are at the core of natural language processing.

Main tasks needed to develop a VPA

There are many tasks involved in processing natural language for a VPA, which can be regrouped in four main categories, described as below.

Speech-to-text and text-to speech

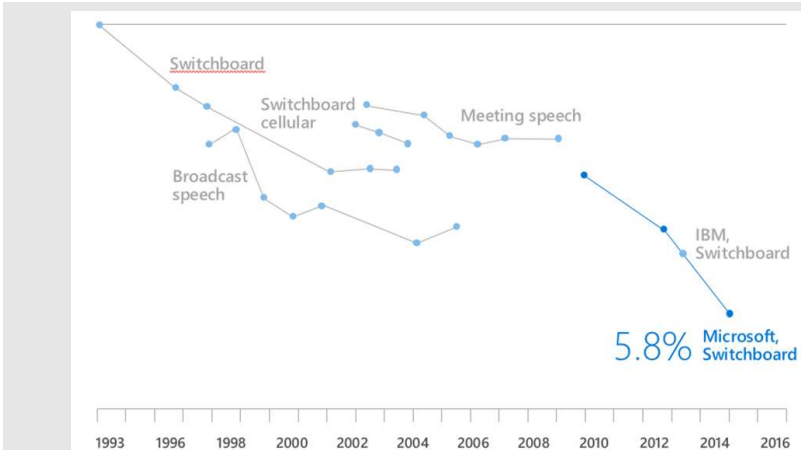
For voice-based input, like those from VPA such as Siri, the first need consists of converting speech to actionable data. This “speech-to-text” step, (also called speech recognition), is of paramount importance: if the input is not correctly recognized, all following steps are useless. Indeed, even an error on one word is very likely to result in an inaccurate answer.

Speech-to-text technologies have been in development for several decades, but the development of AI has enabled important improvements in the last years.

Syntax and semantic processing

Once a sequence of spoken words is successfully converted to a text form, numerous and very complex tasks remain²:

Figure 2: Error rates for speech recognition systems



Notes : According to Microsoft, the error rate for human is 5.9%
Source : Microsoft

- Syntax analysis (or parsing) is used to analyse and identify the structure of the sentence, based on knowledge of grammar.
- Semantic analysis is used to reach a partial representation of the meaning of the sentence, based on the knowledge of the meaning of words.
- Pragmatic analysis is used to reach a final representation of the meaning of the sentence, based on information about the context.

Question Answering

For the vast majority of applications related to natural language, an answer (oral and/or written) is given back after a query from the user.

Using a correctly identified sentence and its meaning/intention, systems thus have to succeed in finding the correct answer and formulating it.

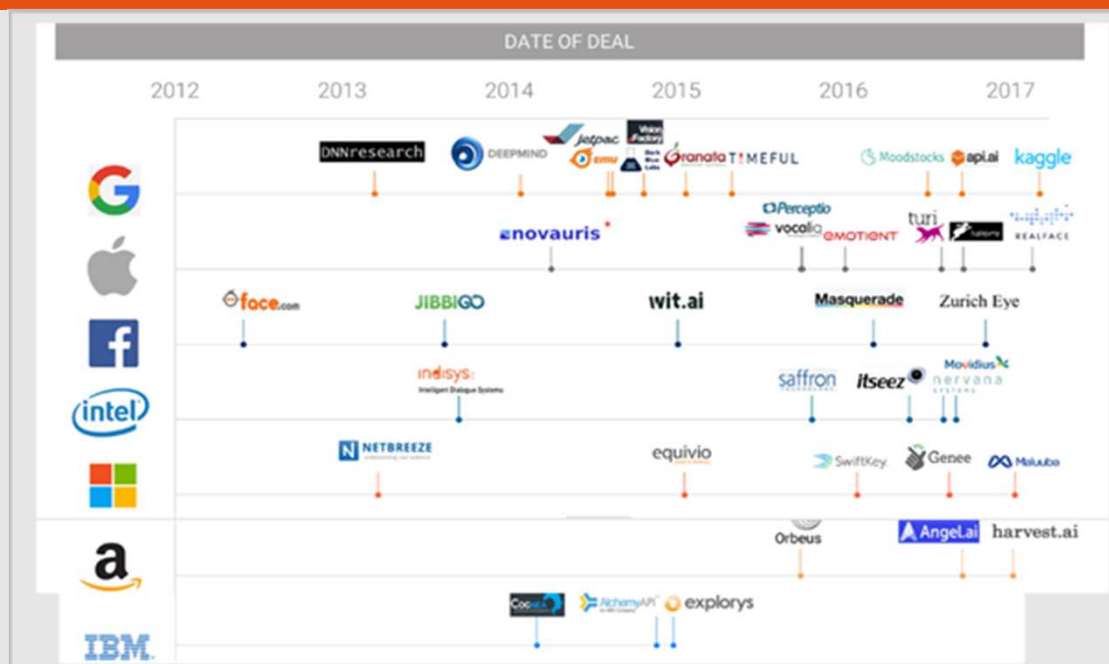
Question answering deals with information retrieval (using information on the Internet, or in an application) and generating a correct sentence, before the last step of speech synthesis.

Text-to speech

Text-to-speech (TTS) is the last step in an VPA interaction which uses audio: the text/answer has already been determined, and the synthesis is the only remaining step.

Speech synthesis is not the most complex task, and is already mastered by all involved players.

Figure 3: Tech Giants takeover of A.I. start-ups



Source: CBInsight, March 2017

Figure 4: Free integration of a VPA, for a third-party developer (Alexa Voice Service)

Natural Voice Control

Alexa has finely tuned automatic speech recognition (ASR) and natural language understanding (NLU) engines that recognize and respond to voice requests instantly. [Learn more about AVS localization](#) »

Always Getting Smarter

Alexa is always getting smarter with new capabilities and services through machine learning, regular API updates, feature launches, and custom skills from the [Alexa Skills Kit \(ASK\)](#).

Free, Easy Integration

The [AVS API](#) is a programming language agnostic service that makes it easy to integrate Alexa into your devices, services, and applications. Best of all, it's free. [Build a prototype with Raspberry Pi](#) »

Source: Amazon

However, one major possibility of improvement is to succeed in generating human-sounding speech. Indeed, current text-to-speech systems are largely based on concatenative TTS and have a robotic-like sound as a result.

2

GAFAM and tech giants are leading the pack

There are many companies, and start-ups, active in the field of AI technologies. However, as GAFAM and tech giants are investing heavily in all AI-related technologies and products, there seem to have an increasing lead regarding VPA.

Currently, these companies are racing to improve the accuracy of their VPA and to add new features, such as:

For the e-commerce giant, securing a central position in the home could be a strategy to develop its prime members' base, and as a result, to further increase its revenue.

- voice recognition: the ability for a VPA to recognize the speaker (in particular for households with more than one person)

- follow-up questions: the ability for a VPA to take into account previous questions and answers

Google

One of Google's most famous public foray into AI technologies is Google Assistant, its VPA available on many Android based products. More generally, the US giant is investing heavily in many artificial intelligence related topics to improve its products:

- Google launched the Cloud Natural Language API in July 2016, a service enabling third-parties to process unstructured data through machine learning.

- In September 2016, Google presented WaveNet a "deep generative model" aiming at significantly improving³ text-to-speech systems by sounding more natural. In October 2017, Google announced that this technology was now used for Google Assistant (for English and Japanese speakers).

- Google funded the creation of a research group to study "deep learning" at the Institut des algorithmes des apprentissages de Montréal (MILA) in November 2016.

Apple

Siri is arguably the most well-known VPA, thanks to an earlier launch than its main competitors, an availability in more languages, and a native support in many Apple devices, iPhones in particular.

Apple has been buying several start-ups in the AI field, in order to improve its products:

- Siri was originally acquired by Apple in 2010, for an estimated 100-250 million USD

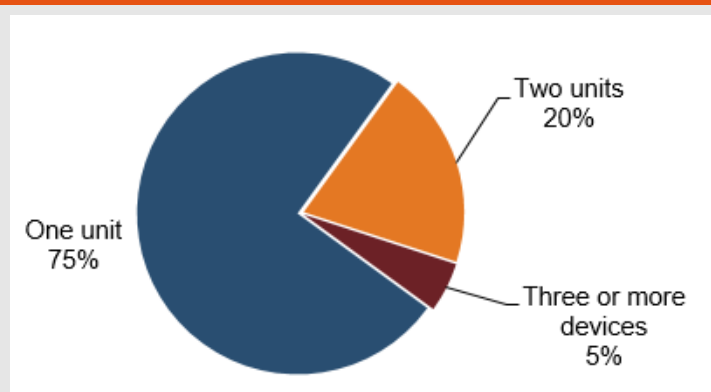
- In 2016, Apple acquired VocallQ, for an estimated 50-100 million USD, with the stated goal to improve Siri.

In June 2016, Apple (partially) opened Siri to app developers, following similar moves from other GAFAM. As a result, Siri can leverage apps from third-party developers and not just Apple's services,

Figure 5: Availability of VPA offered by Internet Giants, by device

Name	Assistant	Launch date	Smartphone tablet	Dedicated device	Computer	TV	Car	Other categories
Amazon	Alexa	2014	✓ (within the Amazon app)	Echo Tap Dot	✗	Fire TV	Hyundai Ford	Plenty
Apple	Siri	2011	✓	HomePod	✓	Apple TV	With CarPlay	Apple Watch
Facebook	M	2015 (still in beta)	✗	✗	✗	✗	✗	✗
Google	Assistant (Now)	2016 (2012)	✓	Home	✓	Android TV	With Android Auto	Android Wear, 3 rd party smart speakers
Microsoft	Cortana	2014	✓	✓	✓	Xbox	Nissan	3 rd party smart speakers

Source: IDATE DigiWorld

Figure 6: Share of households owning multiple VPA-equipped speakers, 2020

Notes: Among households owning at least one VPA-enabled speaker
Source: IDATE DigiWorld, based on forecast data from Gartner

Amazon

Amazon entered the VPA consumer market with the Amazon Echo, a smart speaker using “Alexa” its in-house designed VPA. Since the launch of the Echo in 2014, competitors, notably Google and Apple, launched similar products or are planning to.

Echo has allowed Amazon to establish a presence in the home. In addition to its multimedia and shopping features, Echo can be used as a ‘hub’ to control connected objects (lighting, thermostats, power sockets, etc.) from a multitude of manufacturers. For the e-commerce giant, securing a central position in the home could be a strategy to develop its prime members’ base, and as a result, to further increase its revenue.

The group also set up a 100 million USD fund to help start-ups that want to use its speech recognition technology (Alexa). Amazon is giving access to its AI capabilities, by allowing third-party developers to make their product “Alexa-enabled”.

- Basic compatibility with Alexa (a consumer can control the third-party product from their Echo)
- Full integration (the product is equipped with a speaker and a microphone and acts as an Echo).

3

Key Stakes

For GAFa, AI is a competitive imperative

GAFa are waging multi-sided “wars”, with the goal of trying to impose their platform, ecosystem, products and services.

For instance, Google and Apple are competing on mobile operating systems, a battle that Microsoft as already lost and abandoned.

Processing natural language using technologies based on AI is key for GAFa, as they rely heavily on user input to provide their services. As a result, GAFa have made strong investments in Artificial Intelligence, through acquisitions and partnerships.

The rise of VPA and smart-equipped speakers are arguably the best example of this competition and the tremendous importance of AI. After Amazon’s success in 2014 with its smart speaker Echo, equipped with a VPA, Google and Apple decided quickly to join the race with their own devices, despite the fact that they entered the speech recognition field earlier (with Now and Siri respectively).

This illustrates the fact that these actors want to retain control of their users and do not want to risk that any competitor succeed in developing alone a new type of UI/product. For GAFa, providing a continuity of experience is key in keeping users “locked-in” their platform.

Vertical industries are starting to integrate VPA products

Speech recognition technologies are being integrated into electronic devices in spectacularly rapid fashion.

Initially limited to simple queries, increasingly complex information can now be processed as more and more devices are equipped with an Internet connection and as the field of speech recognition surges forward.

Virtually all consumer electronic (CE) devices can now include speech recognition capabilities:

- “Computers” Smartphones, tablets, computers, smart watches
- “Home entertainment”: TV sets, set-top boxes, streaming boxes, video game consoles
- “Smart home” alarm clocks, speakers, lamps, thermostats, fridges
- Cars

In the vast majority of cases, manufacturers of CE devices with speech recognition capabilities rely directly on the VPA of Amazon, Google or Microsoft. Indeed the three giants allowed third-party developers to fully integrate a VPA (Google Assistant, Cortana and Amazon Alexa respectively) into products.

For vertical industries, there are some risk in integrating a VPA, as they lose a significant portion of their control on the user interface. In addition, the company developing the VPA has a control on which services are available and which are not: they can redirect users towards their own.

Still more U.I. than A.I.

Artificial intelligence technologies have enabled VPA to understand and respond but their promises go beyond being just a new way for end users to interact with their devices (UI). They are seen (and marketed) as intelligent assistants able not just of understanding but of taking decisions, supporting and potentially replacing human in several tasks. This vision has yet to fully materialize.

The long term vision of the development of VPA is that they will become capable of more and more tasks, being for example able to follow entire professional conversations and seek documents in a company information system related to specific requests.

References

- ¹ KPCB - Internet Trends 2016 – Code Conference, July 2016
- ² Worcester Polytechnic Institute - Natural Language Processing - Prof. Carolina Ruiz
- ³ Results (mean opinion scores) available on Deepmind’s website

About the Digital Transformation Monitor

The Digital Transformation Monitor aims to foster the knowledge base on the state of play and evolution of digital transformation in Europe. The site provides a monitoring mechanism to examine key trends in digital transformation. It offers a unique insight into statistics and initiatives to support digital transformation, as well as reports on key industrial and technological opportunities, challenges and policy initiatives related to digital transformation.

Web page: <https://ec.europa.eu/growth/tools-databases/dem/>

This report was prepared for the European Commission, Directorate-General Internal Market, Industry, Entrepreneurship and SMEs; Directorate F: Innovation and Advanced Manufacturing; Unit F/3 KETs, Digital Manufacturing and Interoperability by the consortium composed of PwC, CARSA, IDATE and ESN, under the contract Digital Entrepreneurship Monitor (EASME/COSME/2014/004)

Authors: Vincent Bonneau, IDATE and Laurent Probst, Virginie Lefebvre, PwC

DISCLAIMER – The information and views set out in this publication are those of the author(s) and should not be considered as the official opinions or statements of the European Commission. The Commission does not guarantee the accuracy of the data included in this publication. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use which might be made of the information contained in this publication. © 2017 – European Union. All rights reserved.