

# Table des matières

<b>1</b>	<b>Introduction générale</b>	<b>2</b>
<b>2</b>	<b>Assistants virtuels intelligents</b>	<b>3</b>
2.1	Introduction . . . . .	3
2.2	L'importance du contexte pour un SPA . . . . .	5
2.3	Caractéristiques principales d'un SPA . . . . .	5
2.3.1	Sensible au contexte . . . . .	5
2.3.2	Évolutif . . . . .	6
2.3.3	Multimodal . . . . .	6
2.3.4	Anthropomorphe . . . . .	6
2.3.5	Multi-plateforme et Flexible . . . . .	7
2.4	Domaines d'applications des SPAs . . . . .	7
2.4.1	Vie quotidienne . . . . .	7
2.4.2	Assistance professionnelle . . . . .	7
2.4.3	E-Apprentissage . . . . .	8
2.5	Exemples de SPAs . . . . .	8
2.5.1	Google assistant . . . . .	8
2.5.2	Apple Siri . . . . .	10
2.5.3	Amazon Alexa . . . . .	11
2.5.4	Microsoft Cortana . . . . .	11
2.6	Conclusion . . . . .	11
<b>3</b>	<b>Composants de base d'un assistant personnel</b>	<b>13</b>

To correct  
To explain more deeply

# Chapitre 1

## Introduction générale

- Ici on parlera des motivations qui ont aboutis à ce projet, des objectifs de ce dernier ainsi que ses perspectives

# Chapitre 2

## Assistants virtuels intelligents

### 2.1 Introduction

Depuis la commercialisation du premier ordinateur grand public (Xerox PARC Alto) en 1973, le monde découvrit pour la première fois ce qui allait devenir l'apparence basique de chaque ordinateur moderne. En effet, la compagnie Xerox fut la première à proposer une interface graphique dotée de fenêtres, d'icônes et d'une souris pour se déplacer et d'un clavier pour écrire du texte. Bien que basique, cette idée lança alors plusieurs autres grandes marques sur le même chemin (IBM, Apple, Compaq ...). Par la suite, beaucoup ont essayé d'améliorer la façon dont l'homme utilisait sa machine : souris plus précise, écran doté d'une plus grande résolution, clavier plus enrichi, voire même l'introduction des écrans tactiles dans certains systèmes embarqués.

Cependant, certains voyaient encore cette façon d'utiliser la machine comme trop primitive, et peu intuitive. En effet laissez un enfant devant un ordinateur et il prendrait un bon moment pour apprendre à éditer ne serait ce qu'un simple fichier. Pour citer Donald A. Norman :

“We must design for the way people behave, not for how we would wish them to behave.”[1]

que nous pouvons traduire par :

“Nous devons concevoir selon le comportement des utilisateurs, et non pas selon la façon dont nous voudrions qu'ils se comportent.”

L'humanité a fait beaucoup de chemin depuis les années 70, l'utilisation d'un ordinateur de nos jours avec les moyens classiques (souris, clavier, écran ...) est devenue une tâche triviale, voire même une **seconde nature, cela reste cependant dû au fait que de plus en plus de jeunes enfants sont exposés depuis leur plus jeune âge au monde technologique qui les entoure, le processus d'apprentissage reste cependant présent, l'effort d'utiliser les outils communs reste lui aussi présent.**

La plus naturelle et plus ancienne façon de communiquer pour l'homme a toujours été la parole. Le développement de langues toutes aussi riches et complexes les unes que les autres a permis à l'humanité de briser plusieurs **barrières sociales**. L'avancement le plus naturel pour cette façon de communiquer serait donc de l'étendre aux machines que l'homme a su construire et améliorer au fil des années.

Motivé par cette manière que l'on a de communiquer entre nous, et épaulé par les récentes technologies telles que l'apprentissage automatique, le traitement automatique du langage naturel et

l'intelligence artificielle, les plus brillants des chercheurs ont entamé leurs travaux dans cette toute nouvelle direction.

Les Assistants Virtuels Intelligents (Smart Personal Assistant, SPA [2]) sont donc le produit de plusieurs années de recherche, visant tout d'abord à faciliter certaines tâches pour l'utilisateur. Les premiers SPAs étaient conçus comme des agents de conversation ou Chatbots, limités dans leurs actions et dépendant toujours d'un moyen de communication textuel, ce n'était pas la forme désirée du SPA. Avec l'avancement des recherches sur la reconnaissance automatique de la parole (Automatic Speech Recognition, ASR) et l'émergence de l'apprentissage automatique, les tout premiers assistants virtuels utilisant l'ASR étaient spécialisés dans certains domaines comme des systèmes médicaux d'aide à la décision. Il a ensuite été plus aisé de briser la barrière et de réaliser ce qui était encore une esquisse d'un SPA personnalisé. Aujourd'hui, et ce depuis l'avènement de l'apprentissage profond et la popularisation des Smartphones, de nouveaux SPAs comme Apple Siri (voir 2.5.2) et Google assistant (voir 2.5.1) et Amazon Alexa (voir 2.5.3) ont fait leurs apparitions, offrant de plus en plus de services personnalisés et spécifiques à chaque utilisateurs.

Dans la suite de ce chapitre nous essayerons de mieux détailler ce qu'est un SPA, ce qui est demandé d'un tel système, ses domaines d'application, en enchaînant par une description d'une pseudo-architecture potentielle de ce système, pour enfin conclure sur les limitations actuelles et les motivations de ce projet.

## 2.2 L'importance du contexte pour un SPA

Informellement, un SPA est un type d'agent (voir 2.3.2) logiciel qui peut effectuer certaines tâches et proposer des services dédiés aux utilisateurs qui vont d'une simple tâche (Ouvrir une fenêtre, lancer une application ...) à la réalisation de requêtes un peu plus complexes comme réserver une table dans un restaurant en passant un appel vocal (voir 2.5.1.1). Pour répondre efficacement à toutes sortes de requêtes, un SPA se doit donc de garder trace du contexte courant de sa conversation avec l'utilisateur. Il doit disposer d'un système capable d'enregistrer les informations pertinentes et de savoir les réutiliser, mais aussi de pouvoir déduire lesquelles de ces informations sont manquantes. On parle ici de Context-Awareness ou Sensibilité au contexte, comme vu dans [2].

D'après [2] et [3], *Day* et *Abwod* définissent un contexte comme suit :

“A context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves”

qui peut être traduit par :

“Un contexte est une information qui peut être utilisé pour caractériser l'état d'une entité. Une entité peut être une personne une place ou un objet, considérée comme pertinente à l'interaction entre l'utilisateur et l'application, ainsi qu'à ces deux derniers eux mêmes”

Il en découle que pour parvenir à développer un système qui puisse répondre aux besoins individuels et spécifiques de chaque personne, modéliser et prendre en compte le contexte semble être une solution prometteuse.

## 2.3 Caractéristiques principales d'un SPA

À partir de [2], nous pouvons dégager certaines caractéristiques principales qui peuvent être vues comme primordiales pour qualifier un assistant virtuel comme étant intelligent.

### 2.3.1 Sensible au contexte

Comme précédemment vu dans la définition du contexte (section 2.2), ce dernier peut être interprété comme tout aspect d'une entité (position d'un objet, couleur d'un objet, température d'une chambre, etc.). Un assistant dit intelligent doit donc être capable de capturer le concept du contexte, d'utiliser et de traiter toute information catégorisée comme contextuelle. Pour être plus précis, un SPA doit être sensible à l'évolution du contexte courant, par le biais de capteurs optiques, de microphones, ou tout ce qui pourrait amener l'utilisateur à faire évoluer la requête qu'il a émise. L'assistant devra donc proposer un système de mise à jour du contexte pour éliminer les informations inutiles et garder celles qui pourraient aider à répondre à la requête de l'utilisateur.

### 2.3.2 Évolutif

Comme vu dans la section 2.2, un SPA peut être vu comme un type d'agent. Pour rappel, d'après *Russel* et *Norvig* dans [4], un agent est une entité autonome pouvant interagir avec son environnement afin d'accomplir certaines tâches et peut être de plusieurs types :

- Agent à réflexes simples : agent exécutant ses actions à base de règles conditionnelles simples (c.à.d Si *Condition* alors *exécuter actions*), ils sont ainsi très simplistes et limités dans la portée de leurs actions.
- Agent basé modèle : semblable aux agents à réflexes simples, il est doté d'un modèle interne complexe censé représenter le monde extérieur auquel l'agent a accès. Cependant, il applique les actions de la même manière que le précédent type d'agents.
- Agent à but : ce type représente une amélioration des agents simples puisqu'il est doté d'un ensembles d'états buts à atteindre d'une façon ou d'une autre.
- Agent à utilité : il s'agit ici agents à buts qui tentent d'aboutir à leurs buts d'une manière optimisée (intelligente) utilisant une fonction de mesure adéquate pour le choix des différents états à atteindre.
- Agent apprenant : agent à utilité enrichi par un module d'apprentissage qui sert de juge pour répondre aux "critiques" des actions qu'il entreprend. Le terme agent évolutif est aussi employé.

Pour ce qui est des SPAs, les plus récents systèmes (ex : Amazon Alexa qui améliore son module de reconnaissance de la parole après chaque réponse non *réfutée* par l'utilisateur) peuvent être considérés comme des agents apprenants, répondant de ce fait à la contrainte évolutive imposée. Cependant, le domaine de l'auto-évolution des systèmes intelligents est encore un domaine nouveau qui se voit *aidé* par les récentes avancées dans l'apprentissage automatique [2].

### 2.3.3 Multimodal

Afin d'assurer une aisance d'utilisation, les SPAs sont fréquemment amenés à récupérer les requêtes (ou données) en entrée de la manière la plus naturelle possible (par exemple par le biais de la parole). Cependant, pour garantir une expérience d'utilisation adéquate, l'assistant sera souvent confronté à récupérer ces requêtes de différentes manières, que ce soit à travers une interface graphique (écran tactile) ou à travers un texte brut tapé au clavier, voire même à travers des expressions faciales ou des états cognitives/émotionnels [5], pour ensuite produire une réponse qui elle aussi pourrait éventuellement être de la forme textuelle, sonore ou les deux. Cette capacité à recevoir en entrée et/ou produire une sortie de plusieurs façons différentes est appelée la multi-modalité [6]. Cette caractéristique permet de masquer à l'utilisateur toute la complexité d'acquisition de ses requêtes.

### 2.3.4 Anthropomorphe

Plusieurs auteurs tendent à attribuer une grande importance à l'anthropomorphisme des SPAs [7], qui est

“Un mécanisme qui pousse les êtres humains à induire qu'une entité non-humanoïde possède des caractéristiques et comportements propres à l'homme”[8]

Ce comportement humanoïde pousserait donc l'utilisateur à se sentir plus à l'aise avec l'assistant, le conduisant ainsi à adopter une façon de communiquer plus humaine et moins structurée qu'avec les autres machines. Ceci est une caractéristique majeure d'un SPA se disant personnalisé.

### 2.3.5 Multi-plateforme et Flexible

Malgré leurs récentes prouesses, certains SPAs sont encore restreints à un écosystème fortement dépendant du fabricant. Cowan et al. mentionnent dans [9] que Apple Siri est limité à l'environnement constitué des produits de la firme à la pomme, n'ouvrant par défaut que les applications de cette dernière quand une requête lui est transmise. C'est un comportement que les assistants devraient éviter, car une indépendance des plateformes utilisées est, certes, très complexe à instaurer, mais offre plus de possibilités aux utilisateurs et aux développeurs pouvant ainsi exploiter la puissance de certaines plateformes (Smartphones, TV connectées, etc). Avec l'émergence de l'IoT (Internet of Things) et des maisons intelligentes par exemple, c'est un tout nouveau terrain de jeu qui est présenté aux SPAs, offrant plus d'opportunités pour les utilisateurs.

## 2.4 Domaines d'applications des SPAs

Après avoir vu les différents aspects que les SPAS doivent traiter, nous nous intéresserons maintenant aux types de services et applications que ces derniers pourraient fournir pour démontrer qu'ils peuvent bel et bien faciliter certaines tâches à l'homme.

### 2.4.1 Vie quotidienne

À la base, les SPAs étaient destinés à un usage très personnel comme la gestion des achats dans les supermarchés, ou des guides touristiques de plusieurs destinations de voyage. Cette spécificité a commencé à s'estomper petit à petit avec l'émergence de nouveaux systèmes dédiés à des applications plus générales, comme les maisons intelligentes ou les assistants de planification de tâches. Ceci a permis de mettre encore plus l'accent sur cet aspect de convivialité que les tout premiers SPAs ont tenté de perfectionner. Ainsi, ces assistants spécialisés dans des domaines restreints (Tourisme, shopping, détente, etc) ont été regroupés dans un seul système plus polyvalent, capable de répondre à des besoins quotidiens divers et variés, allant même à fournir une assistance aux personnes âgées pour leur faciliter les tâches rudimentaires devenues trop fatigantes.

### 2.4.2 Assistance professionnelle

Les SPAs ont aussi une place dans le monde professionnel. Dans [10] il est cité que dans les situations où la marge d'erreur est très petite (par exemple dans les système de manufacturing<sup>1</sup>) l'assistance d'un SPA est nécessaire, servant d'un aide à l'humain pour la prise de décision.

Par exemple, dans un environnement de travail hétérogène (Nouveaux/anciens employés, Hiérarchies des postes ...) les SPAs pourraient décharger les employés les plus expérimentés de la tâche

---

1. Manufacturing ici dans le sens chaîne de montage industrielle, par exemple dans des usines.

d'assister les nouveaux arrivants, pour ainsi aider ces derniers dans leurs tâches et permettre aux autres de se focaliser sur les leurs.

### 2.4.3 E-Apprentissage

Les SPAs peuvent aussi être utiles dans l'enseignement, aussi bien dans un milieu académique que professionnel. D'une part ils pourraient occuper plusieurs rôles dans les établissements scolaires (correcteur automatique de copies, enseignant interactif ... ) [11] et, d'autre part, accompagner les employés durant leurs formations professionnelles.

Ainsi, en considérant les caractéristiques d'un SPA, la sensibilité au contexte est reliée aux expériences antérieures de l'apprenant, permettant au SPA d'adapter son processus d'enseignement en conséquence.

En ce qui concerne l'aspect évolutif du SPA, il lui permet de préférer une approche d'enseignement à une autre selon les résultats de ses apprenants.

## 2.5 Exemples de SPAs

Pour illustrer la puissance des SPAs les plus récents, nous présentons dans cette section les quatre produits qui dominent le marché courant :

### 2.5.1 Google assistant

Lancé en 2016 sous forme d'un chatbot intégré dans l'application Google Allo, Google Assistant s'est vu ensuite être directement intégré sur les système d'exploitation Android (que ce soit sur smartphones ou tablettes, et plus récemment sur Google Home<sup>2</sup>). Google Assistant est un assistant à tout faire qui a été développé par les ingénieurs de Google dans le but de faciliter la recherche sur internet, la planification des tâches, l'ajustement des réglages de l'appareil, etc. Son point fort est sa capacité à engager une conversation bi-directionnelle avec l'utilisateur, assurant ainsi une interaction personnalisée variant d'un utilisateur à un autre. Cette capacité lui permet par exemple de proposer certains résultats de recherche selon les précédentes interactions avec l'utilisateur ou de lui proposer une activité si ce dernier lui mentionne qu'il s'ennuie (voir figure 2.1).

---

2. Appareil servant à contrôler les composants d'une smart-house ainsi que l'utilisation des différents services de Google



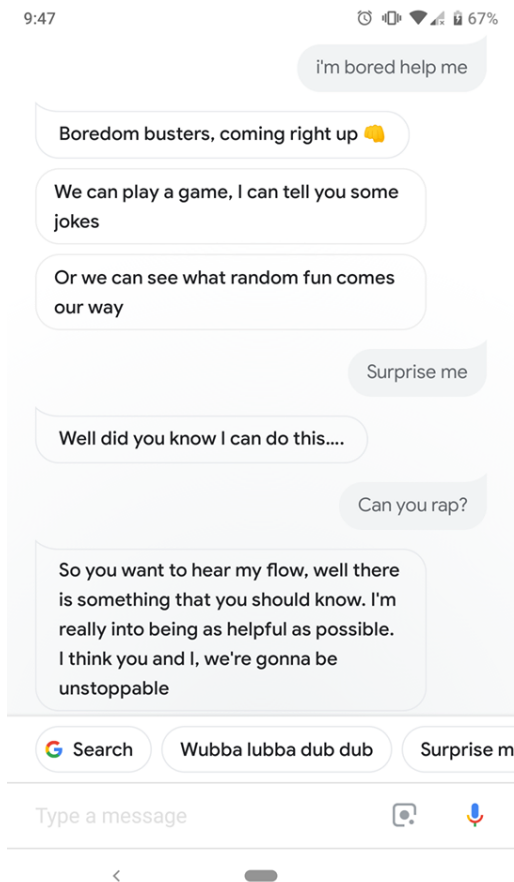


FIGURE 2.1 – *Conversation aléatoire avec Google Assistant*

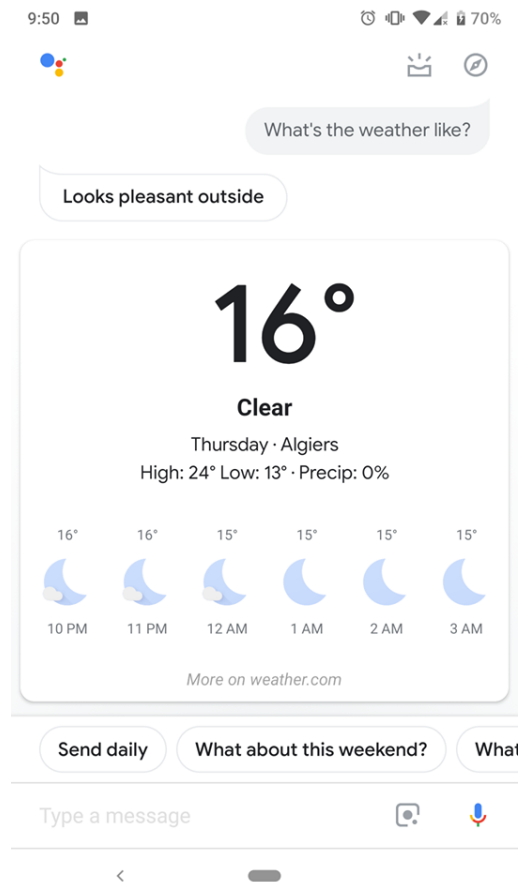


FIGURE 2.2 – *Requête simple formulée à Google Assitant*

### 2.5.1.1 Google duplex

Une des nouveautés impressionnante de Google Assistant est la fonctionnalité Google Duplex. Toujours en phase de développement, ce module est capable de passer des appels a de vraies personnes et d’avoir une conversation avec elles afin de réaliser une tâche demandée par l’utilisateur comme par exemple réserver une chambre d’hôtel, une table au restaurant, etc.

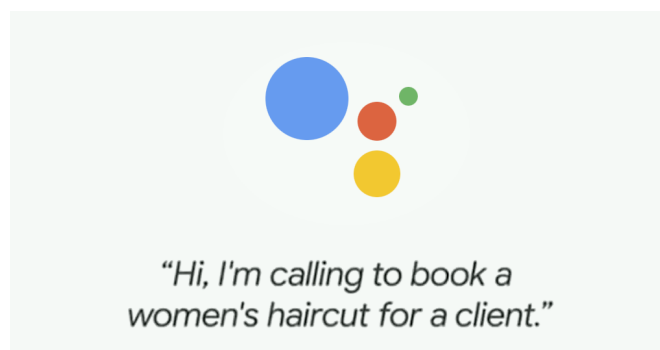


FIGURE 2.3 – *Google duplex réservant une place dans un salon de coiffure*

## 2.5.2 Apple Siri

Siri est l'assistant virtuel développé par Apple. Contrairement aux SPAs durant sa sortie, Siri proposait une nouvelle façon de communiquer avec l'utilisateur, à travers une interface de requêtes vocales, et une façon de **converser** très humanoïde (satisfaisant ainsi le critère d'anthropomorphisme 2.3.4). Siri est capable de répondre à des questions précises (voir figure 2.6), de proposer des recommandations, déléguer la requête à des services web ou d'autres applications (voir figures 2.4 et 2.5). Il a l'avantage (et l'inconvénient) d'être uniquement disponible que sur les appareils qui composent l'écosystème d'Apple (MacBook, iPhone, iWatch, etc).



FIGURE 2.4 – *Intégration aux applications [12]*

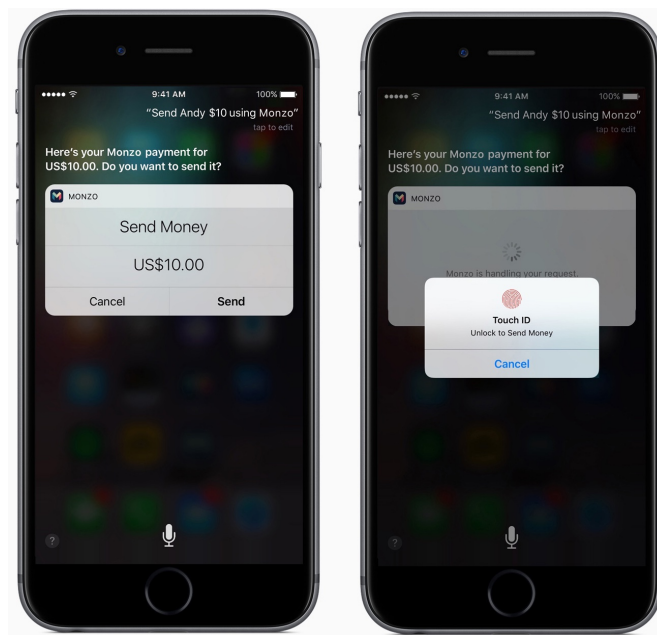


FIGURE 2.5 – *Service paiement 1 [12]*

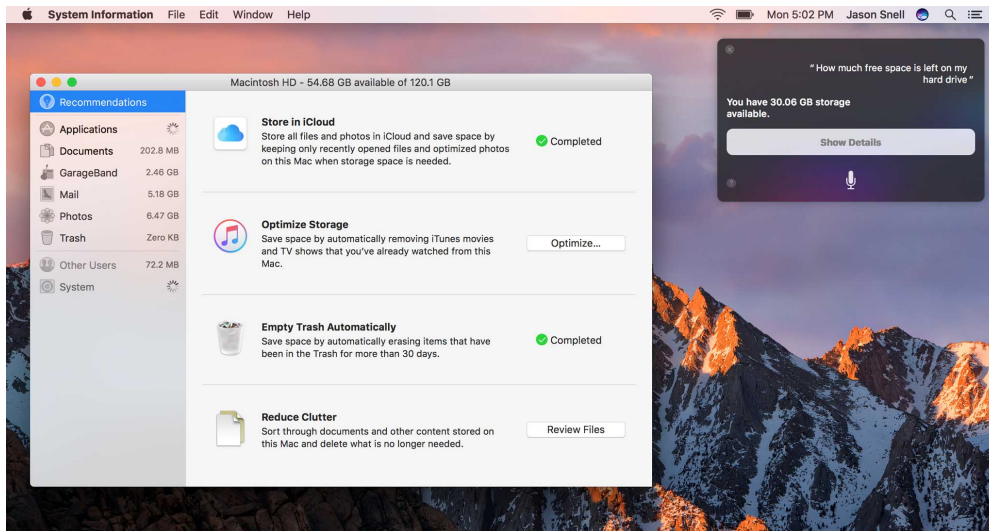


FIGURE 2.6 – Siri sur un laptop [13]

### 2.5.3 Amazon Alexa

Amazon Alexa est un assistant exclusivement intégré au dispositif Amazon Echo (un haut-parleur portatif). À l'instar de Siri, il est aussi capable de communiquer avec l'utilisateur par le biais de la parole, pouvant ainsi exécuter diverse commandes comme jouer de la musique, réciter des livres audios, annoncer des news en temps réel (Résultats sportifs, tendances politiques, etc). Son atout majeur est sa capacité à s'intégrer à plusieurs appareils-connectés (Contrôleur de thermostat ou de lumières ambiantes dans une Smart-House) ainsi que la possibilité d'ajout des Skills (ou compétences) de la part des développeurs tiers pour enrichir la panoplie de services que peut offrir Alexa.

### 2.5.4 Microsoft Cortana

Cortana est la tentative de la part de Microsoft d'intégrer un assistant dans son système d'exploitation Windows 10 et WindowsPhone. Il propose divers services de base tel que planifier des tâches, exécuter des commandes via la parole, et analyser des résultats de recherche sur le moteur de recherche de Microsoft, Bing, pour répondre à des questions.

## 2.6 Conclusion

À travers les sections précédentes, nous avons essayé de présenter les différents aspects d'un assistant virtuel intelligent (caractéristiques, exemples, architectures possibles, etc). Nous avons donc pu apprécier la potentielle puissance d'un tel système s'il venait à être perfectionner d'avantage.

En effet, en examinant les domaines d'applications, il est facile de déduire que le recours à un SPA peut grandement faciliter certaines tâches, que ce soit celles qui sont les plus triviales pouvant retarder d'autres tâches plus importantes, ou bien celles qui doivent faire appel à la précision

et à la grande capacité de calcul des machines, assurant ainsi des résultats précis et rapidement délivrés.

À la fin de ce chapitre nous pouvons donc mettre en valeur la place primordiale que pourraient avoir les SPAs s'ils arrivaient à maturité, c.à.d briser la barrière qui sépare les humains de la machine, parvenant ainsi à faire partie de la vie quotidienne des utilisateurs.

Dans le prochain chapitre nous allons principalement aborder les aspects techniques des différents composants du SPA que nous désirons réaliser.

# Chapitre 3

## Composants de base d'un assistant personnel

### 3.1 Introduction

### 3.2 Architecture d'un SPA

### 3.3 Reconnaissance automatique de la parole (ASR)

#### 3.3.1 Définition

#### 3.3.2 Modèle acoustique

#### 3.3.3 Modèle de langage

#### 3.3.4 Méthodes utilisées

#### 3.3.4.1 N-grammes

#### 3.3.4.2 Modèle de Markov Caché

### 3.4 Compréhension du langage naturel (NLU)

#### 3.4.1 Définition

#### 3.4.2 Classification d'Intent

#### 3.4.3 Extraction d'entités

#### 3.4.4 Analyse sémantique

### 3.5 Gestion du dialogue

#### 3.5.1 Définition

#### 3.5.2 Agent de dialogue

### **3.5.3 Apprentissage par renforcement**

### **3.5.4 Simulateur d'utilisateur**

## **3.6 Génération du langage naturel (NLG)**

### **3.6.1 Définition**

# Table des figures

2.1	Conversation aléatoire avec Google Assistant . . . . .	9
2.2	Requête simple formulée à Google Assitant . . . . .	9
2.3	Google duplex réservant une place dans un salon de coiffure . . . . .	9
2.4	Intégration aux applications [12] . . . . .	10
2.5	Service paiement 1 [12] . . . . .	10
2.6	Siri sur un laptop [13] . . . . .	11



# Bibliographie

- [1] D. A. Norman, *The design of everyday things*. New York : Basic Books, 2002.
- [2] R. Knote, A. Janson, L. Eigenbrod, and M. Söllner, “The what and how of smart personal assistants : Principles and application domains for is research,” in *Multikonferenz Wirtschaftsinformatik (MKWI)*, 2018.
- [3] H. Gellersen, *Handheld and Ubiquitous Computing : First International Symposium, HUC'99, Karlsruhe, Germany, September 27-29, 1999, Proceedings (Lecture Notes in Computer Science)*. Springer, 1999.
- [4] S. J. Russell and P. Norvig, *Artificial Intelligence : A Modern Approach*. Pearson Education, 2 ed., 2003.
- [5] T. Dingler, “Cognition-aware systems as mobile personal assistants,” in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct - UbiComp '16*, ACM Press, 2016.
- [6] E. Luger and A. Sellen, “”like having a really bad PA”,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, ACM Press, 2016.
- [7] R. Trappl, ed., *Your Virtual Butler*. Springer Berlin Heidelberg, 2013.
- [8] A. Purington, J. G. Taft, S. Sannon, N. N. Bazarova, and S. H. Taylor, “”alexa is my new BFF”,” in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17*, ACM Press, 2017.
- [9] B. R. Cowan, N. Pantidi, D. Coyle, K. Morrissey, P. Clarke, S. Al-Shehri, D. Earley, and N. Bandeira, “”what can i help you with?”,” in *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '17*, ACM Press, 2017.
- [10] J. Imtiaz, N. Koch, H. Flatt, J. Jasperneite, M. Voit, and F. van de Camp, “A flexible context-aware assistance system for industrial applications using camera based localization,” in *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)*, IEEE, sep 2014.
- [11] “Engaging the appropriation of technology-mediated learning services – a theory-driven design approach research in progress,” 2015.
- [12] “Apple shares examples of siri’s third-party app integration on ios 10.” <https://www.idownloadblog.com/2016/09/01/apple-siri-ios-10-app-integration/>. (Accessed on 10/29/2018).
- [13] “macos sierra review : Hey siri, where did my files go? - six colors.” <https://sixcolors.com/post/2016/09/sierra-review/>, 2016. (Accessed on 10/29/2018).
- [14] “Understanding the differences between alexa, api.ai, wit.ai, and luis/cortana.” <https://medium.com/@abraham.kang/understanding-the-differences-between-alexa-api-ai-wit-ai-a> (Accessed on 10/29/2018).