

**Faculté des sciences économiques,
sociales, politiques et de communication**

Assistants vocaux : une interaction avec une nouvelle interface homme-machine

**Une approche phénoménographique
centrée sur l'utilisateur**

Auteur : Théo Vaessen
Promoteur : Pierre Fastrez
Année académique 2021-2022
Master 120 en sciences des technologies de
l'information et de la communication à finalité
spécialisée: ingénierie et technologies de la
connaissance

INTRODUCTION.....	4
REVUE DE LA LITTÉRATURE.....	5
1. L'ASSISTANT VOCAL.....	5
1.1 <i>La voix pour interface</i>	6
1.2 <i>Comment ça marche ?</i>	7
2. L'INTERACTION HOMME – AV.....	10
2.1 <i>Une interaction hybride</i>	10
2.2 <i>Détermination des rôles dans l'interaction</i>	13
3. L'UTILISATEUR.....	15
3.1 <i>Usages et contextes d'interaction</i>	16
3.2 <i>La théorie de la domestication</i>	18
3.3 <i>Acceptation de l'assistance vocale</i>	19
3.4 <i>Satisfaction des utilisateurs</i>	20
3.5 <i>Perceptions, évaluations et préoccupations</i>	22
4. LES NOTIONS D'EXPÉRIENCE ET DE COMPÉTENCE APPLIQUÉES AUX AV.....	25
PROBLÉMATIQUE ET MÉTHODOLOGIE.....	27
1. QUESTION DE RECHERCHE.....	27
2. MÉTHODOLOGIE.....	28
3. LA PHÉNOMÉNOGRAPHIE APPLIQUÉE À NOTRE OBJET DE RECHERCHE.....	30
COLLECTE DES DONNÉES.....	31
1. ENTRETIENS.....	31
4.1 <i>Interviews semi directives</i>	31
4.2 <i>Guide d'entretien</i>	32
4.3 <i>Déroulement des entretiens</i>	33
2. RECRUTEMENT ET PROFILS DES PARTICIPANTS.....	34
ANALYSE DES DONNÉES.....	36
1. DÉVELOPPEMENT DE CATÉGORIES.....	38
1.1 <i>Catégories A : Un moyen de divertissement</i>	38
1.2 <i>Catégorie B : Une alternative à l'interaction gestuelle</i>	41
1.3 <i>Catégorie C : Un processus routinier</i>	43
1.4 <i>Catégorie D : Une pratique multitâche</i>	45
1.5 <i>Catégorie E : Un moyen d'organisation et de gestion</i>	47
2. ESPACE DE RÉSULTAT.....	49
2.1 <i>Relations entre les catégories</i>	49
2.2 <i>Organisation hiérarchique</i>	53
3. IDENTIFICATION DE THÈMES.....	54
3.1 <i>Évolution de l'utilisation</i>	56
3.2 <i>Interaction en public</i>	58
3.3 <i>À propos du sentiment de contrôle</i>	60
3.4 <i>À propos de la conversation utilisateur - assistant</i>	62
3.5 <i>À propos de l' (in)compréhension</i>	62
CONCLUSION.....	64
1. IMPLICATIONS POUR LE DÉVELOPPEMENT DES COMPÉTENCES.....	64
2. DISCUSSION.....	66
3. LIMITATIONS ET RECHERCHES ULTÉRIEURES.....	68
4. CONCLUSION GÉNÉRALE.....	71
BIBLIOGRAPHIE.....	72

*Je tiens à remercier tous ceux qui ont été à mes côtés durant cette période
de recherche,*

*Mon promoteur, Pierre Fastrez, pour sa disponibilité, sa bienveillance et son
accompagnement tout au long de ce mémoire.*

*Les personnes charmantes qui se sont portées volontaires pour participer à
mon enquête,*

*Sur le plan personnel, je remercie également mes amis dont la simple
présence me revigore,*

Et enfin ma famille, le pilier de mon épanouissement.

Introduction

Les assistants vocaux (AV) sont des applications qui assistent les utilisateurs dans l'exécution de tâches informatiques. On parle aussi d'assistants virtuels, personnels, numériques ou intelligents. Ces variations terminologiques témoignent de la multitude de variations « conceptuelles » que cette technologie peut susciter. La notion « d'assistance » vocale renvoie à la nouvelle façon (c'est-à-dire l'interaction vocale) d'effectuer des tâches en utilisant des systèmes d'information. La notion « d'intelligence » fait quant à elle référence à l'intelligence artificielle (IA) qui rend possible une forme d'interaction humaine orale avec elle. Le système est effectivement capable de traiter le langage naturel, tout comme il est capable d'en produire.

Dans le cadre de ce mémoire, nous nous intéresserons à la forme la plus largement disponible d'assistance vocale, celle qui est intégrée aux smartphones. Nous pouvons en trouver plusieurs exemples d'implémentation comme Bixby (Samsung), Google Assistant ou encore Siri (Apple). Les concepteurs de ces systèmes exploitent les capacités grandissantes d'interprétation pour tenter de répondre aux besoins dynamiques des utilisateurs actuels. Le présent mémoire est motivé par ce constat et ce qu'il implique, à savoir l'orientation progressive vers ce que certains appellent déjà l'« ère de l'assistance ». Sous l'impulsion du géant américain Google, il y a une volonté croissante d'émancipation des moteurs de recherche traditionnels, au profit d'une assistance capable de proposer une réponse appropriée, directe et rapide à l'utilisateur (Tisserand-Barthole, 2017).

Plusieurs chercheurs s'accordent pour dire que les assistants vocaux ont le pouvoir de changer la façon dont les utilisateurs consomment l'information (Cowan et al., 2017). Par conséquent, les utilisateurs sont les premiers acteurs concernés et nous nous sommes dès lors interrogés sur leur compréhension du système et leur expérience d'interaction avec lui. Notre question de recherche est la suivante : « Quelles sont les différentes façons

qualitatives dont les utilisateurs comprennent l'interaction avec leur assistant vocal ? ».

Plusieurs études ont déjà abordé la question des assistants vocaux par le prisme de l'utilisateur. Notre étude aura comme spécificité de faire état des différentes conceptions des utilisateurs à propos de l'interaction avec l'AV et de tenter d'ordonner celles-ci. Pour répondre à ces besoins, une étude phénoménographique sera menée. Nous débuterons ce mémoire par une revue de la littérature, avant de nous intéresser à notre problématique et à la description de notre recherche empirique. Enfin, après avoir présenté et analysé nos données, nous terminerons par une conclusion transversale, dans laquelle nous discuterons des résultats de notre mémoire et de leurs implications pour d'éventuelles exploitations ultérieures.

Revue de la littérature

1. L'assistant vocal

La popularité de l'assistance vocale s'est développée parallèlement à l'essor des smartphones, et, plus récemment, des réseaux de téléphonie mobile. En effet, l'assistant vocal requiert le plus souvent une connexion internet, du fait de sa dépendance à des serveurs distants.¹ Au-delà de l'intelligence artificielle dont cette technologie profite, la particularité de l'assistant vocal réside dans sa capacité à construire de nouvelles connaissances à partir de son expérience (Fréjus & Lahoual, 2018). L'expérience dont il est question consiste à recueillir des données vocales fournies par l'utilisateur. Ces données, toujours plus nombreuses à l'ère du Big Data, permettent l'apprentissage automatique de la machine. On parle « d'algorithmes d'apprentissage machine ». Cependant, nous le verrons, le chemin est encore long pour que les AV deviennent d'excellents assistants. Pour Budju & Whitenton (2018) comme pour d'autres auteurs (Budju &

¹ cfr. section 1.2 Comment ça marche ?

Laubheimer, 2018) il existe de grands écarts entre les besoins des utilisateurs, la manière dont ils exploitent le système et le potentiel des AV d'aujourd'hui.

1.1 La voix pour interface

Pour décrire l'assistant vocal, commençons par dire qu'il ne se contente pas d'exécuter des tâches, il propose aussi une réponse en langage naturel et peut avoir recours à des sources d'information externes (telles que l'agenda ou le courrier électronique) pour déduire les actions potentielles qui intéressent l'utilisateur.

Pour transmettre les informations et permettre l'interaction avec l'utilisateur, tout l'enjeu réside dans la qualité de l'interface. Pour l'AV, on parle parfois d'interface « naturelle ». Celle-ci fait référence à l'aspect invisible de l'interface utilisateur. Dans cette perspective, l'affordance² devrait être intuitive ou transparente. C'est ce que suggère Suchman (2007) : l'objectif de l'utilisabilité est un « artefact auto-explicatif ». Autrement dit, pour qu'un consommateur adopte une technologie, son fonctionnement doit être apparent. Cependant, certaines études (Turunen et al., 2009 ; DeRenzi et al., 2017) ont souligné le désir de complémentarité prôné par bon nombre d'utilisateurs entre la modalité vocale et d'autres modalités (notamment tactile et visuelle) pour faciliter l'utilisation.

L'interaction avec les agents « conversationnels » pose des questions depuis leur création. En tant que technologie mobilisant le langage naturel, les assistants vocaux peuvent donner l'illusion d'interactions conversationnelles avec l'utilisateur. De nombreuses recherches ont cependant mis en avant les différences fondamentales avec les interactions humaines. Ces divergences majeures s'appuient notamment sur le caractère asymétrique de l'interaction homme-machine (Suchman, 2007). Malgré les

² Terme popularisé par D. Norman, couramment utilisé dans le domaine de la conception d'interface centrée utilisateurs (UX), et qui se définit comme « la capacité d'un objet ou d'un système à évoquer sa propre utilisation, sa propre fonction. » (Alain, 2019)

efforts en termes de « machine learning », la compréhension du contexte interactionnel est limité pour l'assistant vocal, notamment en ce qui concerne l'interprétation des requêtes ou l'incapacité à détecter les troubles de la conversation (Hayes & Reddy, 1983). Les réponses produites par l'AV sont programmées selon un principe de commande à 2 tours : l'utilisateur émet une requête et le système y répond. Suivant cette logique, on comprend qu'il est compliqué pour les AV de répondre de manière « situé » à une stimulation de la part de l'utilisateur. Cependant, des avancées récentes dans ce domaine, en particulier pour l'assistant de Google, rendent possible l'intégration d'un certain nombre d'éléments de contexte. Pour rendre compte des progrès dans ce domaine, nous proposons un exemple concret d'utilisation avec Google Assistant (LaterClips, 2019) :

- Interlocuteur : « À quelle heure arrive ton vol ? »

L'assistant interprète le message et affiche automatiquement les détails du vol que l'utilisateur a réservé en ligne. Sur base de ces informations, l'utilisateur s'adresse à l'assistant :

- Utilisateur : « Réponds que j'arrive à 15h14. »
- L'assistant répond à l'interlocuteur : « J'arrive à 15h14. »

Dans cet exemple, l'AV connaît le contexte. Sans que l'utilisateur n'ait à préciser à qui il désire adresser son message, l'assistant comprend qu'il doit répondre à un message récent contenant le mot « vol ». L'utilisateur n'a pas à « ré-entreprendre » un processus à deux tours en précisant certaines informations (message exacte, destinataire, confirmation d'envoi...) (LaterClips, 2019).

1.2 Comment ça marche ?

Sans rentrer dans trop de détails techniques, il nous semble important de planter les bases du fonctionnement d'un assistant vocal. Ces quelques éléments de compréhension nous permettront plus tard de mettre en perspective les expériences d'interaction évoquées par les utilisateurs.

La conception et le fonctionnement des assistants virtuels dit « génériques » (Siri, Google, Alexas, ect.) se divisent en plusieurs modules. Comme nous l’avons mentionné, ces modules dérivent principalement d’une discipline : le traitement automatique du langage naturel. Parmi les modules embarqués, on peut citer la reconnaissance de la parole, le traitement sémantique (hors et en contexte), la gestion de l’historique et des stratégies du dialogue, l’accès aux ontologies, la gestion des accès aux connaissances externes (base de données ou internet) et la génération de langage (synthèse de la parole) (CNPEN³, 2020).

Pour y voir plus clair, il est nécessaire de donner quelques précisions sur l’architecture du système de l’assistant vocal. Les principaux assistants personnels intelligents utilisent le processus décrit par Bellegarda dans la figure 1.

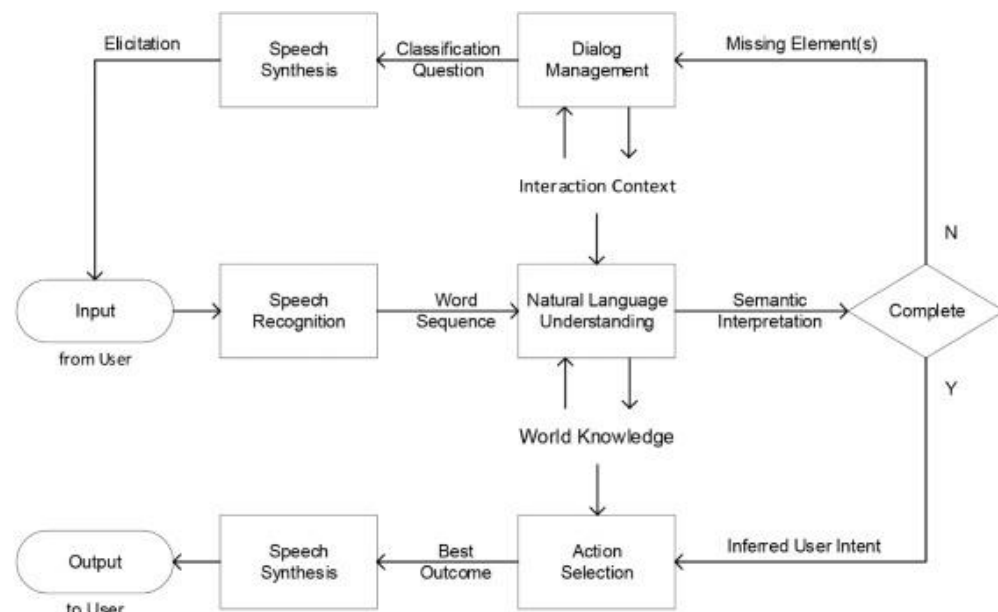


Figure 1 : « Intelligent Personal Assistant » Interaction Model (Bellegarda, 2013).

³ Comité National Pilote D’éthique du Numérique

L'énoncé vocal est d'abord transcrit en une séquence de mots. C'est ce qui permet la compréhension du langage naturel, la langue « ordinaire » parlée par un être humain. Le sens des mots est ensuite interprété par le système. Si ce dernier manque d'éléments de compréhension, il passe par la « gestion du dialogue » qui s'appuie sur le contexte d'interaction pour obtenir des informations supplémentaires. Une fois l'interprétation sémantique terminée, la connaissance de la tâche guide la sélection de la meilleure action pour la situation en question. Enfin, le résultat sélectionné est transmis à l'utilisateur. La CNIL⁴ a publié une infographie plus récente qui vient compléter le modèle d'interaction de Bellegarda :

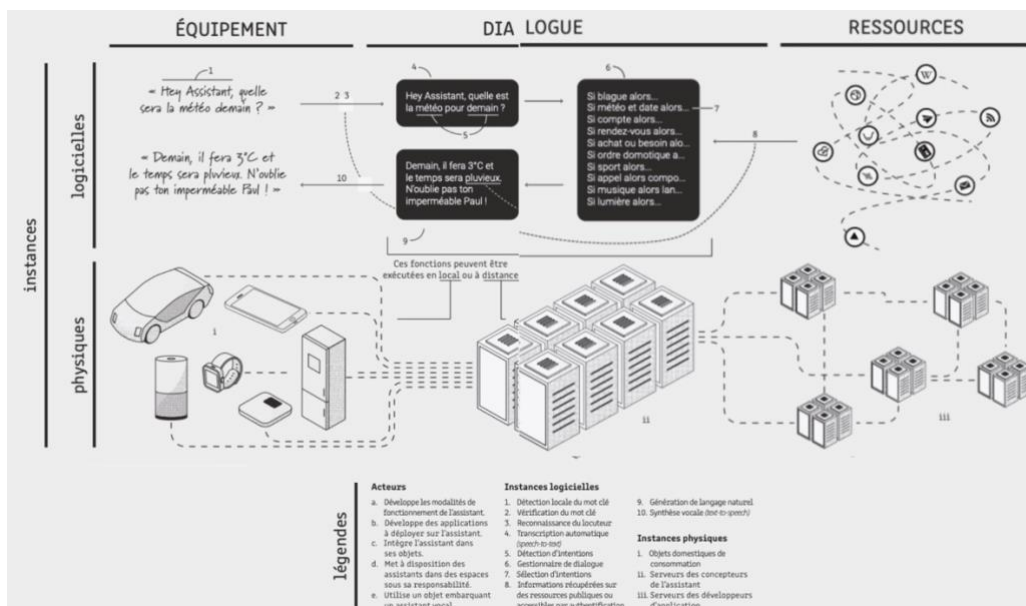


Figure 2 : La voie des données (CNIL,2020).

La CNIL précise notamment que pour que l'interaction soit possible, la plupart des assistants vocaux doivent désormais être « réveillés ». Autrement dit, l'assistant doit basculer dans un mode d'écoute actif pour recevoir les ordres et commandes de son utilisateur. Ce réveil peut être réalisé suite à une action physique, mais le plus souvent il se fait sur base de la détection d'un mot-clé (on parle de mot d'activation ou en anglais wake-up word / hot word). Quand une interaction prend fin, l'assistant vocal repasse

⁴ Commission Nationale de l'Informatique et des Libertés

en veille. Par ailleurs, pour fournir la réponse adaptée à la requête de l'utilisateur, des ressources distantes peuvent être mobilisées. Cela peut être des connaissances publiquement accessibles (encyclopédie en ligne, etc.) ou des informations accessibles par authentification (compte bancaire, application musicale, compte client pour achat en ligne, etc.). Ces informations peuvent être issues des serveurs des concepteurs de l'assistant ou de serveurs des développeurs d'application. La capacité de connexions à de telles sources de données est une caractéristique essentielle de l'assistant virtuel basé sur le cloud.

En effet, ces serveurs auxquels on accède par Internet, ainsi que les logiciels et les bases de données qui fonctionnent sur ces serveurs, constituent ce qu'on appelle le « cloud computing ». Les informations traitées par l'assistant vocal ne sont pas stockées directement sur le smartphone. Cela permet d'économiser de l'espace et de stocker davantage de données. En outre, cela permet à l'appareil de réagir automatiquement lorsqu'il est appelé. Le système transmet en continu ce que le microphone capte à un service de cloud, qui lui-même transmet des informations en retour. En l'absence de cette technologie, le système d'assistance s'est longtemps confronté à un problème d'instantanéité. Il était impensable d'avoir un agent conversationnel fluide, car un décalage entre les commandes et les réponses aurait eu un impact considérable sur l'expérience utilisateur (Katarey, 2020).

2. L'interaction Homme – AV

2.1 Une interaction hybride

Comme nous l'avons évoqué, les utilisateurs utilisent le langage naturel pour interagir avec l'AV. En outre, ils mobilisent des pratiques « ordinaires », c'est-à-dire des structures de phrase comparables à celles qu'ils utilisent dans la vie de tous les jours avec leurs congénères. Plusieurs auteurs ont donc employé la notion « d'hybridité » (Krummheuer, 2015 ; Relieu & Velkovska, 2020 ; Zouinar, 2019) pour qualifier l'interaction avec l'AV, qui se situe à mi-chemin entre la conversation et l'interaction homme-machine.

Dans cette perspective, Laubheimer et Budiu (2018) suggèrent que les utilisateurs peuvent être classés en 2 catégories ; ceux qui utilisent la même structure de langage que l'homme, et ceux qui utilisent d'autres artifices pour tenter d'être mieux compris (usage de mots clés, de commandes, d'impératifs, etc.).

En ce qui concerne la première catégorie, des auteurs comme Qvarfordt, Jönsson et Dahlbäck (2003) ont remarqué très tôt la tendance naturelle à personnifier l'assistant. On attribue volontiers des attributs humains à la machine. Ce constat ne date donc pas d'hier, des études le démontrent depuis plusieurs décennies, et cela ne constitue pas toujours un acte délibéré de la part de l'utilisateur (Nass et al., 1994). Dans le cas des assistants vocaux, l'anthropomorphisation se manifeste surtout à travers l'usage de formules de politesse (« bonjour », « merci », « s'il te plaît », etc.) ou des questions appelant à la « réflexion » de l'appareil (« qu'en penses-tu ? », « pourquoi ? », « et toi ? » etc.). Dans une logique marchande, les fabricants jouent sur cette idée de personnification en proposant des assistants emphatiques et « gentils ». Les concepteurs leur confèrent souvent des traits d'humour ou des préférences divers pour faciliter et fortifier l'adhésion de l'utilisateur. L'assistant, auquel on attribue un prénom (Alexa, Siri, etc.) et une voix (généralement féminine), devient ainsi un objet social. Il est présenté comme le parfait bras droit pour l'utilisateur, à la fois disponible, rassurant et compétant. L'engagement de l'utilisateur et la confiance dans les informations présentées sont ainsi renforcés. De cette personnification émane un questionnement sur le caractère social de l'assistant vocal. Plusieurs études se sont penchées sur ces questions par le biais des usages et de leur contexte.

Une étude récente a démontré que les utilisateurs avaient principalement recours à leur assistant vocal pour l'accomplissement d'un nombre relativement restreint d'activités simples (Budju et Laubheimer, 2018). Cette observation met en lumière un paradoxe lié à cette technologie relativement nouvelle : de plus en plus de gens l'adoptent malgré la faible « usabilité ». L'assistant vocal n'est donc pas utilisé dans tous les cas où il pourrait l'être.

Selon le Nielsen Norman Group, seuls 6,8% des besoins adressés par l'utilisateur se feraient via l'assistant vocal. Les gestes sur les supports physiques et en particulier les ordinateurs ou les smartphones sont encore largement priorités.

Cela s'explique en partie par le fait que pour qu'un assistant puisse répondre à un besoin avec succès, deux conditions préalables doivent être remplies (Budju et Laubheimer, 2018.) :

- 1) le participant doit avoir le bon assistant (car tous les assistants ne se valent pas pour toutes les tâches)
- 2) le participant doit formuler la commande adéquate pour ce besoin.

Budju et Laubheimer concluent dans leur étude de 2018 menée sur « l'usabilité » des assistants vocaux que l'utilisateur attend finalement peu de ces outils. Cela se traduit par le fait qu'il les réserve à des questions factuelles et relativement binaires. L'utilisateur ne privilégie la modalité vocale que pour un nombre restreint de ses activités quotidiennes. En effet, d'autres supports (donc d'autres modalités) sont encore bien plus ancrés dans les mœurs. Ainsi, malgré la popularisation manifeste de la technologie, un utilisateur disposant d'un assistant vocal et d'un ordinateur capable d'effectuer une tâche similaire, et dans une circonstance pour laquelle il est en pleine possession de ses moyens, n'optera généralement pas pour l'assistant. Nous prendrons cependant le soin de nuancer cette thèse en nous appuyant sur les résultats de notre étude empirique.

En guise d'amorce, rappelons que dresser une opposition entre un assistant vocal, un smartphone et un ordinateur n'a pas beaucoup de sens, du fait que l'assistant vocal est incarné dans une diversité d'artefacts, regroupant notamment l'ordinateur et le smartphone. En fonction des supports physiques à sa disposition, l'utilisateur évalue quelles sont les modalités (vocales, gestuelles ou tactiles) qui correspondent le mieux à la fonction (la tâche) qu'il cherche à effectuer. Cet effort peut s'avérer laborieux pour l'utilisateur car le fait de faire « correspondre » une modalité et une tâche renvoie à des éléments difficilement tangibles. Parmi ces éléments on peut évoquer la commodité de

l'usage, la qualité ou la complétude du résultat escompté, ou encore le niveau de « contrôle » que l'utilisateur désire, à plusieurs égards.

Dans la plupart des cas, on constate que les interactions entre les utilisateurs et les systèmes d'assistants vocaux sont plus complexes que les recherches sur Internet (Burbach et al., 2019). Par exemple, les systèmes d'assistants vocaux doivent être capables de comprendre l'intention et le contexte de l'utilisateur afin de pouvoir choisir l'action appropriée ou fournir une réponse adéquate. On peut faire l'hypothèse qu'il s'agisse d'un facteur d'explication de l'utilisation limitée de l'assistant vocal. Dans son expérience, l'usager aurait tendance à s'orienter vers une solution qui limite les éventuelles incompréhensions, or la réalisation d'une tâche par un geste sur les supports physiques existant offre davantage de contrôle pour l'utilisateur. Nous spécifierons la nature des obstacles rencontrés par les usagers dans l'adoption de l'assistance vocale dans les sections relatives à l'acceptation et la satisfaction.

2.2 Détermination des rôles dans l'interaction

Un élément fondamental de la communication est de savoir « qui » participe à l'interaction. Dans le cadre d'une interaction avec un AV, l'utilisateur et l'assistant semblent être les seules parties à échanger des messages. Guzman (2017) parle d'une relation interpersonnelle dyadique. Dans cette relation, l'assistant travaille pour l'utilisateur et est contrôlé par ce dernier. Il est cependant nécessaire de distinguer l'interlocuteur du support. L'utilisateur s'adresse à son assistant (son interlocuteur), puis sa commande se voit prise en compte et exécutée sur un smartphone ou sur un autre outil (le support). En d'autres termes, l'AV nous parle en tant qu'interlocuteur et travaille pour nous en tant que médium (Guzman, 2017). Au moins 3 entités sont donc impliquées dans cette interaction, ce qui la distingue ici encore de la relation humaine. En fait, l'assistant est un logiciel qui reçoit et échange des informations non seulement avec l'usager, mais aussi avec d'autres supports et avec toute société ou organisation impliquée dans le processus d'interaction décrit plus haut.

Dans une recherche sur l'interaction homme-machine, Valérie Saint-Dizier de Almeida (2009) s'est penchée sur la dimension relationnelle de la communication. Dans cette étude, l'auteure met en place une méthode pour recueillir des données destinées à la conception d'un système d'assistance interactif dans un domaine spécifique (la géologie). Dans son étude, seules deux entités sont considérées dans l'interaction, mais pour la comprendre et la différencier d'une interaction conversationnelle traditionnelle, sa méthode consistait à faire croire à un sujet qu'il interagissait avec une machine, alors qu'il interagissait en réalité avec un humain (un expert) chargé de simuler la machine. Ainsi, le statut des « partenaires » de l'interaction change. L'auteure parle d'une interaction « a-didactique », qui se distingue de l'interaction « didactique » du fait d'un changement de paradigme intentionnel. Dans le premier cas, celui qui nous intéresse, l'enjeu est la résolution d'une tâche (relation résolveur-assistant) alors que dans le second cas l'intention est celle de l'apprentissage (relation maître-élève). C'est à partir de cette démarche que Saint-Dizier de Almeida détermine les rôles de l'interaction entre un assistant (« compères » de l'étude) et un utilisateur (« sujet » de l'étude). Nous spécifions dans quelles mesures les rôles identifiés sont pertinents dans le cadre de l'interaction avec les AV génériques que nous avons l'habitude de manipuler aujourd'hui :

Les assistants :

- a. exécutent les commandes (actions simples, paramétrage, création et gestion d'évènement dans l'agenda, contenu de divertissement, etc.)
- b. décrivent l'état en cours (ou l'état de la tâche : « opération en cours », « opération effectuée »,...). Cette catégorie est moins transposable aux assistants intelligents actuels, pour lesquels les réponses sont souvent instantanées. Il est néanmoins fréquent qu'un pictogramme ou qu'une animation visuelle remplace cette fonction.
- c. expliquent (notamment les raisons pour lesquelles une commande ne peut être exécutée)
- d. définissent (communiquent des définitions ou des informations)

- e. assistent (participent à rendre exécutables des commandes). La rédaction de messages textes par dictée vocale est une incarnation parmi d'autres de ce rôle.
- f. informent (informent sur la terminologie du domaine de l'application - le lexique à employer pour la formulation des commandes). On peut faire un parallèle avec les « suggestions d'utilisation » qui sont généralement proposées à l'ouverture des AV, et qui visent à donner aux utilisateurs une meilleure visibilité sur les capacités du système et sur les formulations adéquates.
- g. guident (des interventions qui permettent de poursuivre la résolution d'une tâche lorsque le sujet ne sait plus quoi faire). Ce rôle existe, notamment dans les interactions à plusieurs tours, mais reste assez peu courant.

Les utilisateurs :

- a) ordonnent (les actions logicielles : les commandes)
- b) informent (donne des informations sur les attributs de commandes)
L'utilisateur peut vouloir spécifier ses commandes et donner davantage d'informations qu'avec la modalité gestuelle, par exemple.
- c) acceptent (les suggestions sont acceptées (ou pas))
- d) questionnent (demande de définition ou demande d'information sur la procédure). Avec la consigne « que peux-tu faire ? », l'utilisateur accède généralement à une liste de suggestions de la part de l'AV.
- e) remercient. Rôle qui fait écho à l'idée d'anthropomorphisme expliqué précédemment. On peut y ajouter toutes autres formes de personnification.

3. L'utilisateur

La littérature scientifique considère encore globalement les utilisateurs d'assistants vocaux comme des premiers adhérents (Trudeau-Pepin, 2020). Ce terme (dérivé de l'anglais : « early adopters ») désigne, selon Rogers dans sa théorie sur la diffusion de l'innovation (Rogers, 1995), la minorité de

technophiles qualifiée qui constitue la première vague d'adoption d'une technologie. Ces adhérents précoces représenteraient 13,5 % de la population. Il s'agit d'un public qui se caractérise généralement par une attitude plus tolérante vis-à-vis de la technologie, lorsque celle-ci n'est pas complètement aboutie et qu'elle présente encore des erreurs. Cette catégorie d'utilisateur joue aussi un rôle « d'influenceur » et contribue à la popularisation d'une technologie au sein du reste de la population. Cependant, nous pensons que ces caractéristiques sont trop génériques, et qu'elles ne sont pas suffisamment satisfaisantes pour décrire les utilisateurs actuels d'assistants vocaux.

3.1 Usages et contextes d'interaction

Au niveau des études sur l'usage des assistants vocaux, plusieurs chercheurs nous éclairent précisément sur les contextes d'utilisation. Budju et Laubheimer (2018) ont observé les contextes d'utilisation les plus courants sur un échantillon d'utilisateurs. Il en ressort que l'AV est employée prioritairement dans deux situations distinctes. La première lorsque les mains sont occupées et la seconde lorsque poser une question est plus rapide que de la taper et de lire les résultats. Une autre étude précise que la première catégorie regrouperait une majorité d'utilisateurs ayant recours à l'AV en conduisant (Heitzman, 2019). Pour la deuxième catégorie, Budju et Laubheimer spécifient que les questions concernées sont généralement des questions simples, succinctes et précises. Les questions ouvertes et nuancées devront souvent être simplifiées par l'utilisateur pour être traitées correctement par l'AV. Dans une autre étude, les mêmes auteurs (Budju et Laubheimer, 2018) ont évalué les AV selon 6 techniques d'interface utilisateur : l'entrée vocale, le langage naturel, la sortie vocale, l'interprétation intelligente, l'intégration et « l'entremise ». Il en résulte que l'AV ne satisfait aucun de ces critères à l'heure actuelle. En effet, ils ne répondent qu'aux exigences minimales d'utilisabilité pour les interactions simples. L'interprétation de phrases complexes est en cause, tout comme les réponses vocales qui ne correspondent pas systématiquement aux questions posées.

D'autres études nous éclairent sur les types de tâches communément adressées sur l'AV. D'une manière générale, l'utilisation la plus courante semble être la recherche d'informations simples (définitions, résultats sportifs, mesures, géographie, etc.). Viennent ensuite la consultation de la météo et la communication avec un interlocuteur (appel, SMS ou e-mail) (Kathryn Whintont, 2018). Dans son étude, Kathryn Whintont va plus loin et décrit quatre catégories de tâches, en fonction de leur complexité :

- 1) Les actions simples ne nécessitant qu'une seule étape ou une action simple pour être réalisées (minuteur, météo, définition, etc.).
- 2) Les tâches « multi-étapes » qui nécessitent de passer par plusieurs étapes pour achever un processus. (appeler un Uber, passer une commande en ligne etc.)
- 3) Les activités multitâches qui impliquent plusieurs activités et/ou applications pour atteindre un objectif. (création d'une liste à partir d'informations fragmentées, opérations de sauvegarde, etc.)
- 4) Les activités de recherche qui nécessitent de rassembler plusieurs sources d'information et d'analyser les options (trouver le meilleur hôtel sur base d'un ensemble de critères).

La grande majorité des utilisateurs se limiteraient à la première catégorie. La deuxième est adressée dans une moindre mesure, et les deux dernières, qui correspondent à des niveaux de complexité élevés, ne le seraient pratiquement jamais. Par ailleurs, les auteurs de l'étude précisent que les tâches adressées par les utilisateurs sont souvent répétées (par exemple la consultation de la météo tous les matins) et que cela conditionne certainement la conception de l'interaction avec l'AV. Ces propos sont corroborés par Nir Eyal, spécialiste des thèmes à l'intersection entre la psychologie et la technologie. Ce dernier soutient l'idée selon laquelle l'usage des AV est particulièrement stéréotypé et comportemental. En effet, même si la typologie des usages reste relativement simpliste, ces objets ont pour fonction de modifier nos habitudes (Gershgorn, 2018). C'est un enjeu majeur pour les fabricants de faire en sorte

que l'utilisateur se tourne de lui-même vers l'AV pour un nombre grandissant de tâches.

3.2 La théorie de la domestication

Nous pensons qu'il est important de compléter le point précédent sur l'usage de la technologie avec une approche anthropocentrée. Ainsi, nous expliquons brièvement ici la théorie de la domestication, proposée par Roger Silverstone (2005). Ce chercheur s'est penché sur la façon dont un média rencontre son public. Concrètement, il a cherché à savoir si « la technologie était utilisée de la façon prescrite par les concepteurs, ou si elle était façonnée pour répondre aux besoins particuliers de l'utilisateur » (Trudeau-Pepin, 2020). Le processus dit de « domestication » comprend quatre phases: l'appropriation, l'objectification, l'intégration et la conversion. Chacune de ces phases a une influence sur l'usage concret, et le processus, bien qu'il ne s'enchaîne pas nécessairement de manière linéaire, démontre que les usages et leur évaluations évoluent dans le temps (Fréjus & Lahoual, 2018).

- 1) La phase d'appropriation est celle qui suit l'acquisition de la technologie par l'utilisateur. On constate parfois rapidement un décalage entre les usages observés lors de cette phase et les usages préconisés par les concepteurs.
- 2) La phase d'objectification considère l'aspect « spatial » lorsque l'objet est acquis. En l'occurrence, l'objectification de l'AV tel qu'on l'entend dans notre étude correspond à peu de chose près à celle du smartphone, bien que le contexte dans lequel on l'utilise est différent.
- 3) La troisième phase, ou « phase d'incorporation », s'intéresse à l'utilisation réelle du dispositif. Concrètement, il s'agit de la phase durant laquelle la technologie s'introduit dans le quotidien de l'utilisateur.

- 4) Enfin, la phase de conversion s'intéresse aux relations entre l'utilisateur, la technologie et le monde extérieur. C'est le moment où l'expérience d'utilisation rencontre la sphère publique. Cela passe aussi par le partage d'expérience de l'utilisateur.

Les quatre phases de ce processus devrait nous aider à mieux contextualiser l'expérience d'utilisation de nos participants dans la partie empirique de notre étude. La vérification de la pertinence de ces différentes phases dans l'appropriation de l'assistance vocale sortirait de notre objet de recherche, mais pensons pouvoir nous appuyer sur cette notion théorique pour établir des ponts entre l'expérience d'interaction et l'appropriation concrète.

3.3 Acceptation de l'assistance vocale

Nous avons déjà mentionné que l'utilisateur ne s'appuie pas autant qu'il le pourrait sur son assistant vocal. Pour tenter de comprendre ce phénomène, on peut l'appliquer au modèle d'acceptation de la technologie (TAM en anglais) proposé très tôt (1989) par le chercheur Fred Davis . Ce modèle spécifie les relations causales entre les caractéristiques de conception du système, l'utilité perçue, la facilité d'utilisation perçue, l'attitude envers l'utilisation et le comportement d'utilisation réel. Davis constate ainsi que l'utilité perçue a une influence 50 % plus importante que la facilité d'utilisation pour déterminer l'utilisation, ce qui souligne l'importance d'incorporer les capacités fonctionnelles appropriées dans les nouveaux systèmes (Davis, 1993). Plus récemment, Yang & Lee (2019) ont constaté que l'utilité perçue et le plaisir perçu ont des effets significatifs sur les intentions d'utilisation en ce qui concerne les assistants vocaux. Pour aller plus loin et aborder d'autres facteurs, le modèle de Davis a été étendu récemment dans le but de comprendre l'intention d'adoption et d'utilisation d'un assistant virtuel d'intelligence artificielle (Song, 2019). Sa recherche a confirmé que l'utilité perçue a des effets positifs significatifs sur l'attitude envers un assistant virtuel et sur l'intention comportementale de l'utiliser. À l'inverse, si l'utilisateur projette sur l'assistant l'idée d'un outil « gadget », n'offrant pratiquement aucune plus-value par rapport aux supports qu'il a

l'habitude de manipuler, il n'en exploitera probablement pas tout le potentiel. De plus, il a démontré que la facilité d'utilisation perçue avait des effets positifs significatifs sur l'utilité perçue, l'attitude envers un assistant virtuel et l'intention comportementale de l'utiliser.

En termes d'acceptation de la technologie, d'autres facteurs plus « actuels », tel que la confiance, la sécurité ou encore la confidentialité perçues par les individus sont également à prendre en compte. Ces facteurs⁵ peuvent influencer sur l'adoption et l'utilisation. Sur ces questions, selon une enquête de PricewaterhouseCoopers (2018), 38 % des non-utilisateurs d'assistants virtuels ont mentionné qu'ils avaient refusé d'adopter la technologie parce qu'ils ne voulaient pas d'un appareil surveillant leur vie quotidienne, et 28 % ont précisé que c'était en raison de leur malaise face aux questions de sécurité des données et de confidentialité. Lau Zimmerman et Schaub (2018) affirment quant à eux que les utilisateurs d'AV sont beaucoup moins sceptiques par rapport aux respects des données et à la vie privée que les non-utilisateurs.

3.4 Satisfaction des utilisateurs

La satisfaction résultant d'une interaction avec un assistant vocal se distingue de celle résultant d'une recherche sur le web, de par la nature de ces interactions. En effet, l'interaction avec l'assistant vocal embarque des dimensions supplémentaires :

- a) la reconnaissance automatique de la parole (ASR). Il a été démontré que les erreurs de reconnaissance vocale influencent négativement la satisfaction de l'utilisateur (Jiang et al., 2013) ;
- b) la compréhension de l'intention de l'utilisateur ;
- c) une interaction basée sur le dialogue (maintien du contexte) ;

⁵ Nous revenons brièvement sur ces facteurs dans la section suivante : 3.3 Perceptions, évaluations et préoccupations

- d) le besoin d'information complexes : les utilisateurs expriment des besoins d'information plus sophistiqués lorsqu'ils interagissent avec des assistants intelligents. (Kiseleva et al., 2016)

Ces dimensions supplémentaires permettent différents scénarios d'utilisation (contrôle de l'appareil, recherche d'informations, recherche embarquant des sous-tâches, etc.) pour lesquels la satisfaction des utilisateurs varie. Kiseleva et ses collègues (2016) ont mis en avant que la nature et la complexité de la tâche à accomplir influent sur la notion de satisfaction. En outre, ils ont démontré que la préservation du contexte de la conversation (dont les progrès récents ont été exemplifiés dans le point 1.1) est essentielle, et que la satisfaction globale de la tâche ne peut être réduite à la seule satisfaction de la requête. Ils concluent que l'achèvement de la tâche est fortement corrélé à la satisfaction de l'utilisateur dans le cas du scénario de contrôle de l'appareil - soit il a fonctionné, soit il n'a pas fonctionné. Pour les scénarios de recherche d'informations, la satisfaction de l'utilisateur est davantage liée à l'effort qu'à l'achèvement de la tâche (Kiseleva et al., 2016). Par ailleurs, dans une étude portant sur les interactions avec Alexa (l'assistant de l'entreprise Amazon), les utilisateurs ont déclaré être satisfaits de l'assistant même lorsqu'il ne produisait pas les informations recherchées, ce qui suggère que l'expérience d'interaction est plus importante pour les utilisateurs que le résultat de l'interaction (Lopatovska et al., 2019).

Bellegarda (2013) précise quant à lui que le « succès » de l'interaction se mesure en termes subjectifs. Tout dépendra des besoins de l'utilisateur par rapport à son intention et à ses attentes. En fonction de la tâche, une interaction « réussie » peut se traduire par une interaction "efficace" (avec un minimum d'interruption), "minutieuse" (la tâche est vraiment terminée) et/ou "agréable" (comparable à une interaction avec un assistant humain).

3.5 Perceptions, évaluations et préoccupations

Plusieurs chercheurs en IHM (interaction homme-machine) se sont déjà intéressés aux attentes, aux perceptions et aux expériences des utilisateurs par rapport aux AV, ou plus généralement aux interfaces vocales. Les études s'accordent à dire que les utilisateurs ont tendance à utiliser le dialogue humain comme métaphore de l'interaction, les attentes en matière de compétence et de fonctionnement du système étant déterminées par la ressemblance humaine de leur conception (Yılmaz & Rızvanoğlu, 2021). Les modèles mentaux générés par les utilisateurs sur cette base sont erronés et créent des attentes irréalistes par rapport aux capacités du système (Luger & Sellen, 2016). Par conséquent, les utilisateurs auraient beaucoup de mal à évaluer le niveau d'intelligence du système, ce qui constituerait un frein pour effectuer des tâches plus complexes. Pour comprendre ce concept de « gouffre d'évaluation », on peut se référer au modèle de l'interaction de Norman. Pour ce dernier, le concept d'interaction entre un utilisateur et un dispositif (objet, ordinateur, etc.) se divise en 4 parties (Norman, 1990) :

- a) But : l'idée que l'utilisateur se fait de ce qui devrait arriver avec le dispositif.
- b) Exécution : l'action que l'utilisateur décide d'effectuer.
- c) Monde : l'utilisateur effectue des actions sur le monde (actions physiques sur le dispositif).
- d) Évaluation : l'utilisateur valide l'action effectuée en comparant le résultat de l'action sur le monde par rapport à son but.

Ces 4 parties constituent un cycle d'exécution et d'évaluation de l'action et l'accomplissement d'une tâche par un utilisateur se traduit généralement par plusieurs itérations de ce cycle.

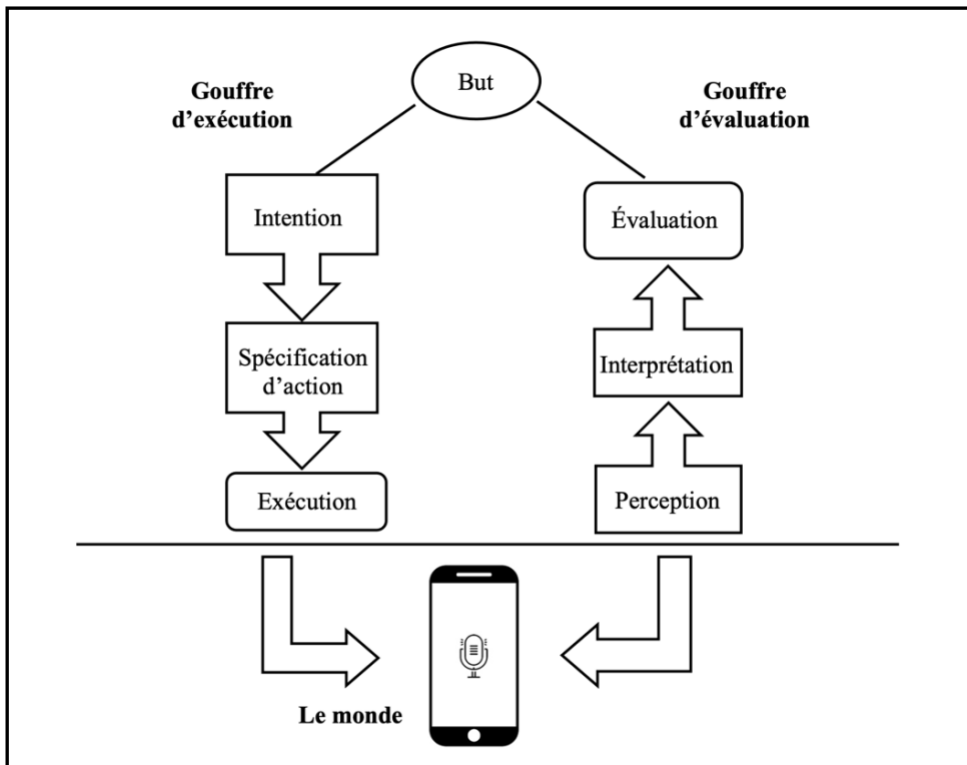


Figure 3 : cycle d'exécution/ évaluation de l'action selon Norman

Cependant, l'utilisateur peut rencontrer des problèmes dans l'accomplissement du cycle, soit au niveau de l'évaluation, soit au niveau de l'exécution. Norman parle d'un « gouffre d'exécution » lorsque le modèle mental que s'est fait l'utilisateur quant aux actions à accomplir pour effectuer une tâche diffère des actions réelles à accomplir (« interactions Visuelles Numérique », 2013). Cela entraîne généralement l'absence d'action, ou l'exécution d'actions inutiles ou infructueuses. Par exemple, si l'utilisateur a l'habitude de s'exprimer d'une certaine manière avec son AV afin d'obtenir une information, mais qu'au cours d'une interaction le système interactif n'est pas en mesure de le comprendre, l'utilisateur devra alors interrompre le processus cyclique et réévaluer ses actions. Comme évoqué plus haut, l'utilisateur peut également se confronter à un « gouffre d'évaluation ». S'il n'est pas capable de « localiser » son prochain objectif dans l'interface, ou qu'il n'arrive pas à évaluer de manière claire quel a été l'effet de son action sur le système, il ne sera pas en mesure de continuer son interaction (« interactions Visuelles Numérique », 2013). Par exemple, si les « tâches » prises en charge par l'AV ne sont pas clairement définies, l'utilisateur devra

instinctivement tenter des commandes, sans avoir l'assurance d'obtenir satisfaction (atteindre le « but » de la tâche). Ces problèmes d'interactions décrits par Norman ont une influence sur l'effort cognitif et l'expérience utilisateur. Le « gouffre d'évaluation » entraînerait également un désir de la part d'une majorité d'utilisateur de confirmation visuelle des activités entreprises avec l'assistant (Luger & Sellen, 2016).

Une autre étude a tenté de nuancer les problèmes de perception en comparant des utilisateurs fréquents et des utilisateurs occasionnels (Yılmaz & Rızvanoğlu, 2021). Certains problèmes communs sont exprimés par les deux types d'utilisateurs, comme le manque de confiance dans l'exécution des tâches par l'AV (qui persisterait donc dans le temps) ou les difficultés rencontrées en situation d'interaction de type « main-libre ». Par contre, les inconvénients potentiels évoqués plus haut concernant l'homme comme référent métaphorique de l'interaction semble être un obstacle plus important pour les utilisateurs occasionnels que pour les utilisateurs fréquents (R. Cowan et al., 2017).

Au niveau des perceptions et des préoccupations des utilisateurs par rapport aux AV, une autre question centrale est celle de l'enregistrement, du traitement et du stockage de leurs données. Sans rentrer dans le détail des recherches dans le domaine de la confidentialité, plusieurs études ont permis de constater une certaine défiance de la part des utilisateurs, du fait de l'intrusion que l'AV représente dans leur vie privée. Ce scepticisme à l'égard de l'AV peut constituer un frein à son utilisation (Budju & Laubheimer, 2018). Il peut s'exprimer à plusieurs niveaux, qui correspondent aussi à des niveaux de compréhension du système. Ainsi, les préoccupations majeures concerneraient l'enregistrement de données (y compris contre le gré de l'utilisateur), leur transmission sur le cloud et le traitement de celles-ci, mais aussi la gêne sociale que peut constituer le fait de parler en public. Nogueira et ses collègues (2017) ont mené une étude qui indique qu'environ 81,8 % des utilisateurs sont d'accord pour dire que les méthodes actuelles seraient moins invasives s'il existait une législation protégeant la confidentialité des données partagées. De même, les utilisateurs seraient favorables à un contrôle

plus précis des informations partagées. En effet, les principaux fabricants commerciaux s'appuient sur des plateformes (IOS, Windows, Android, etc.) qui sont autorisées à lire les données de l'utilisateur, mais l'utilisateur n'est généralement pas conscient du type de données que l'application consomme.

En bref, la nature même de la technologie explique en partie les difficultés rencontrées par les usagers dans la perception et l'expérience avec elle. Les mêmes raisons qui confèrent à l'assistant une dimension « magique », du fait de l'opacité du mécanisme, sont aussi à la source des difficultés de compréhension. Chun (2011) va dans le sens de cette réflexion en affirmant que le logiciel est difficile à comprendre parce qu'il est éphémère. Nous ne voyons pas comment s'opère le processus, et nous nous appuyons plutôt sur la métaphore du logiciel pour construire notre compréhension de l'informatique. L'interface vocale laisse plus de place à l'imagination qu'une interface graphique traditionnelle. Cela explique en partie les projections métaphoriques de la part des usagers, et laisse davantage de place à l'élaboration de fantasmes en termes de confidentialité (si tant est qu'ils s'agissent de fantasmes).

4. Les notions d'expérience et de compétence appliquées aux AV

L'évolution technologique est fulgurante dans le domaine de la communication. Cette évolution s'accompagne de nouveaux services et de produits qui se basent sur un certain nombre de connaissances (Sandberg, 2000). La gestion et la combinaison de ces connaissances constituent les compétences de l'utilisateur. Pour Roegiers, dans Scallon (2004), la compétence serait « la possibilité, pour un individu, de mobiliser de manière intériorisée un ensemble intégré de ressources en vue de résoudre une famille de situations-problèmes ». Un élément constitutif de la compétence seraient donc les connaissances organisées pour construire une représentation cohérente et fonctionnelle d'un système d'action.

Jean-Claude Coule (2010) affirme quant à lui que l'expérience du sujet peut, d'une part, nourrir l'activité qu'il déploie en situation, tout en se construisant à travers cette activité. Suivant cette logique, dans le domaine managérial, on n'hésite pas à mettre en place des formations pour la gestion et le développement des compétences, ce qui assurent le succès compétitifs (Kanter, 1983 ; Porter, 1990).

Pour décrire les pratiques d'un sujet , on aura donc souvent tendance à penser que son activité est en grande partie dictée par son apprentissage et ses acquis. Le fait d'être compétent avec un outil ne va pas sans une certaine expérience, de même que quelqu'un d'expérimenté est généralement perçu, à priori, comme compétent. Dans le domaine des TIC, cela fait écho à l'approche dite « technocentrée », qui soutient l'idée selon laquelle l'homme doit s'adapter à une technologie et non le contraire (Brousseau & Rallet, 1997). L'utilisateur doit donc développer des connaissances et des compétences pour utiliser correctement la technologie.

Cette approche n'est cependant pas la seule qui existe, et elle est contestée en ce qui concerne le cas des assistants vocaux. Shulevitz (2018) ainsi que Lemaire et ses collègues (2019) estiment que le cas des AV respecte davantage l'approche « anthropocentrée », car ils s'intègrent rapidement dans la routine de l'utilisateur. Il nous apparaît donc nécessaire de revoir la manière avec laquelle s'articulent les notions de compétence et d'expérience, particulièrement dans la compréhension globale des interactions avec les AV.

Ceci étant dit, nous précisons que les objectifs de cette recherche ne sont pas celle d'un manager. Pour nous, comprendre comment les utilisateurs font l'expérience de l'interaction avec un AV, c'est comprendre l'influence de leur niveau de compréhension sur leur usage et sur les compétences observables. Le « manager » pourra ensuite, par exemple, élaborer des démarches de formation ou d'éducation afin de développer les compétences, si compétences il y a. Dans notre étude, nous ne cherchons pas à faire état d'un nombre fini de connaissances et d'aptitudes mais de comprendre quelles sont celles que

les usagers mobilisent et conçoivent dans le cadre de leur activité avec l'assistant vocal.

Problématique et méthodologie

1. Question de recherche

Comme nous l'avons mentionné, l'assistance vocale s'impose de plus en plus dans notre vie quotidienne (smartphones, voitures, domotique, etc.) mais nombreux sont ceux qui s'accordent pour dire que la recherche académique peine encore à développer des méthodes pour étudier cette technologie (Pailleur et al., 2020). Le développement massif de ce marché soulève pourtant des questions importantes sur les formes d'interaction et d'appropriation, sur la transformation des activités et des relations sociales, ou encore sur les conséquences sur la vie privée (Velkovska et al., 2020).

Dans la recherche qui nous concerne, l'utilisateur est au centre du jeu. C'est son expérience de l'interaction avec l'assistant qui nous intéresse. En effet, nous ne chercherons pas particulièrement à étudier la technologie en elle-même, mais plutôt les variations dans la façon dont son usage est vécu, afin de comprendre notamment comment s'articulent les notions de compétence et d'expérience. Autrement dit, l'objectif visé est la compréhension interprétative des interactions avec un assistant vocal, à partir des perspectives expérientielles des utilisateurs. Notre question de recherche est la suivante : « Quelles sont les différentes façons qualitatives dont les utilisateurs comprennent l'interaction avec leur assistant vocal ? »

La façon de poser notre question de recherche, et l'hypothèse selon laquelle l'interaction avec les assistants vocaux fait l'objet de conceptions différentes en fonction des utilisateurs, émane d'une théorie, la « phénoménographie », que nous décrivons dans la section suivante. Par le biais de cette méthode, nous chercherons également à comprendre et à ordonner les différents niveaux de compréhension identifiés. Notons

cependant qu'une conception plus complexe/complète ne doit pas nécessairement générer un usage plus complet : il doit permettre à l'utilisateur de développer l'usage qui lui convient (et qui peut être limité, pour telle ou telle raison, tout en reposant sur une conception complexe).

Conjointement à cet effort de catégorisation, nous nous interrogerons également sur les concepts et les expériences communes exprimées par les utilisateurs dans leur relation avec leur AV. Dans un espace de discussion, nous tenterons enfin d'émettre quelques hypothèses pour d'éventuelles recherches ultérieures par rapport aux variables telles que l'âge ou le secteur d'activité, et la manière avec laquelle elles affectent potentiellement la relation avec les AV.

2. Méthodologie

Plusieurs théories et processus associés répondent à l'exigence de comparaison des expériences des utilisateurs. Celle qui fait office de référence pour nous s'appelle la "phénoménographie". Cette théorie, et le processus auquel elle est liée, consiste à se focaliser sur les conceptions qu'ont les gens d'un phénomène et à décrire ces dernières dans le but de comprendre comment ledit phénomène peut-être vécu. Marton (1986) et Booth (1997), ont décrit la phénoménographie comme suit :

"Phenomenography is focused on the ways of experiencing different phenomena, ways of seeing them, knowing about them and having skills related to them. The aim is, however, not to find the singular essence, but the variation and the architecture of this variation by different aspects that define the phenomena".

La phénoménographie recherche la variation plutôt que les points communs des expériences d'un phénomène. Il s'agit donc de considérer une « situation » à partir de plusieurs angles différents. Seul un regard holistique et complet permet d'incorporer la variation dans l'analyse de l'expérience.

En outre, en comprend que la méthode de recherche phénoménographique est ancrée dans la réalité perçue plutôt que dans les cadres théoriques préconçus des chercheurs. En ce sens, elle est particulièrement intéressante pour surmonter les idées préconçues autour de l'appropriation et de l'usage d'une technologie. Dans la littérature existante, assez peu de recherches similaires ont été menées dans le domaine des TIC. Bien que la phénoménographie s'intéresse à un ensemble très hétérogène de phénomènes, certains domaines semblent avoir été davantage investigués. C'est notamment le cas de « l'information literacy » ou les sciences de l'éducation.

Dans ces domaines comme dans d'autres, l'approche phénoménographique a permis de remettre en question des concepts communément admis. Dans le domaine de l'apprentissage par exemple, les travaux menés par Ference Marton ont remis en question l'idée selon laquelle la réussite de l'apprentissage peut être mesurée en évaluant le degré de correspondance réelle entre la forme originale de l'information présentée à l'étudiant et la forme de l'information telle qu'elle est présentée par l'étudiant dans les conditions de l'examen, en partant du principe que cette forme « d'apprentissage » (appelée apprentissage de surface par Marton) entrave en fait une compréhension plus profonde. Un exemple parmi d'autres qui nous permet de mieux saisir l'intérêt de l'approche phénoménographique. Pour comprendre ce que signifie le concept d'« apprentissage » et ce qu'il implique, il est nécessaire de confronter les différentes conceptions existantes autour de celui-ci. Il en va de même pour tous les phénomènes du monde qui nous entoure. Si on se limite à une conception communément admise, on ne peut difficilement tirer des conclusions satisfaisantes sur le phénomène étudié. Ceci dit, la définition de la « compréhension » d'un phénomène est à la fois très vague et très complexe. Kennedy (1978) le formule comme suit :

“Although current models and theories about the processes of comprehension are partial, discrete, and, as yet, inadequate, a variety of approaches to the processes involved have revealed the active constructive nature of comprehension on the part of the listener, both

at the levels of speech perception and interpretation. [...] We perceive partly what we expect to hear, and what we want to hear as well as what is said, and these are related to highly individual background experiences.”

Kennedy mentionne ici l’aspect profondément individuel de l’expérience dans le processus de compréhension. Afin d’apporter quelques détails sur ce point et d’approfondir la compréhension de notre cadre théorique, nous allons nous référer spécifiquement à notre objet de recherche dans la section suivante.

3. La phénoménographie appliquée à notre objet de recherche

Dans le cas de l’utilisation d’un assistant personnel, l’utilisateur interagit effectivement seul avec le dispositif. Dans le processus interactionnel, on peut considérer l’utilisateur comme étant le sujet et l’assistant comme étant l’objet de notre l’étude. Notre place (celle du chercheur) se situe entre ces deux entités. Pour bien comprendre la position du phénoménographe, on peut la mettre en perspective avec celle des approches constructivistes « internes » et « sociales ». Dans la première, les actes et les comportements externes sont le reflet des actes mentaux « internes ». Autrement dit, les réalités sont construites par les individus. Dans la seconde approche, ce sont les actes et les comportements « externes » qui expliquent les actes mentaux « internes ». L’environnement impose ses réalités à l’individu (Abeysekera, 2018). La phénoménographie refuse cette distinction marquée entre un monde « interne » et un monde « externe ». Elle privilégie une réalité émergente des interactions entre l’environnement et l’individu.

En d’autres termes, la phénoménographie examine le phénomène de recherche à partir d’une perspective de « second ordre », se concentrant sur les façons dont les personnes qui vivent le phénomène en font l’expérience (Marton, 1997 ; Limberg, 2000 ; Andretta, 2012). L’enjeu est ensuite d’établir une variation dans les manières collectives dont ces phénomènes sont vécus. Marton et ses compères mettent donc l’accent sur la multitude de façons de vivre un phénomène dans la pratique. Il va de soi qu’établir la variation des

conceptions d'un phénomène n'est possible qu'en prenant en compte l'addition de chacune d'elles. Ainsi, bien que la phénoménographie soit "sensible à l'individualité des conceptions" des phénomènes de recherche (Ashworth & Lucas, 2000), l'approche phénoménographique mettra davantage l'accent sur la compréhension « collective » d'un phénomène plutôt que sur l'expérience individuelle.

Enfin, Limberg (2000) affirme que des recherches phénoménographiques ont « *confirmé l'idée que la manière dont les gens font face à un problème ou agissent dans une situation est liée à la manière dont ils vivent ces problèmes ou ces situations. La capacité d'agir d'une certaine manière reflète la capacité de faire l'expérience de quelque chose d'une certaine manière.* » Suivant ces propos, nous pouvons partir du principe que les conceptions des utilisateurs, ou leurs façons de comprendre l'interaction, et leur utilisation de l'assistant virtuel sont interdépendants. Dans cette optique, l'interaction peut être conceptualisée comme ayant à la fois un "comment" et un "quoi". Les actions constitutives de l'interaction ne seront pas séparées du concept personnel d'interaction.

Collecte des données

1. Entretiens

4.1 Interviews semi directives

Nos participants ont été interrogés selon l'interview semi-directive. Cette technique consiste à interroger les utilisateurs suivant un ensemble restreint de questions bien précises. Contrairement à l'interview directive, l'ordre des questions prévues peut varier légèrement d'une interview à l'autre, favorisant la fluidité de l'échange et la mise en confiance des participants. En tant qu'intervieweur, il est primordial de garder le contrôle de l'entretien, en recentrant les propos en cas de digression d'une part, et en « canalisant » les réponses conformément aux objectifs fixés d'autre part.

En amont du premier entretien, nous avons fixé notre objectif : nous informer sur l'expérience passée et actuelle de l'interaction avec l'AV, mais aussi et surtout soulever des connaissances implicites inscrites dans l'expérience faites avec lui. Conformément à l'approche phénoménographique, nos questions incitent le participant à mettre au jour sa compréhension conceptuelle de l'interaction. L'enjeu est en effet d'amener nos sujet à mettre des mots sur le déroulement d'actions qui sont souvent inconscientes. Par conséquent, nous choisissons d'axer globalement nos consignes sur le « comment » et le « quoi » plutôt que sur le « pourquoi » des actions. Nous cherchons à décrire davantage la manière avec laquelle l'interaction est vécue plutôt que l'origine intentionnelle. Sans négliger pour autant les causes et les raisons des processus, nous mettrons donc l'accent sur l'explication de ces derniers.

4.2 Guide d'entretien

Pour constituer notre questionnaire (cfr. annexe 2) nous nous sommes appuyés sur la revue de littérature réalisée dans la première moitié de ce travail de recherche. L'utilisation de travaux et de recherches antérieurs sur notre sujet constitue pour nous, au même titre que la théorie liée à l'approche phénoménographique, une base solide permettant de connaître les expériences antérieures "pertinentes". Marton (1994) affirme en effet que le phénoménographe doit, préférentiellement, se baser sur un support de présuppositions, pour permettre d'éviter tout jugement. Nous avons organisé notre guide selon les différentes dimensions de notre recherche (fonction, satisfaction, incompréhension, dimension temporelle, procédé, définition générale, attentes). L'ordre des questions, auquel nous conférons une légère flexibilité (conformément à ce qui est admis dans le cadre d'entretien semi-directif), a été conçu pour passer de l'expérience personnelle à des généralités abordées sur la base d'expériences propres.

Pour ce qui est de l'expérience personnelle, la première dimension, relative à la « fonction », s'intéresse à l'utilisation concrète de l'utilisateur (à quoi lui sert l'assistant ?) et aux circonstances d'interaction (moments, lieux,

situations). Nous faisons ensuite appel à une expérience d'interaction récente et concrète, dont la description par l'utilisateur doit nous permettre de saisir les éléments de satisfaction et/ou de contrariété rencontrés. Nous nous penchons ensuite sur la question de l'incompréhension dans l'interaction, et la justification que peuvent en faire nos usagers. Ensuite vient la dimension temporelle, pour laquelle nous cherchons à savoir si l'utilisation passée de notre utilisateur a permis d'ajuster son utilisation actuelle. De cette manière, nous tentons d'évaluer à quel point les obstacles persistent dans l'expérience d'interaction. La dimension suivante s'intéresse à ce que l'utilisateur comprend du procédé d'interaction, en demandant à ce celui-ci de décrire l'usage de l'AV à un non-initié. Enfin, nous sortons de l'expérience purement personnelle, en invitant nos répondants à donner une définition de l'assistant vocal et à se projeter dans le futur pour en proposer une hypothétique représentation.

4.3 Déroulement des entretiens

Notre questionnaire de recrutement nous a permis d'aborder nos interviews avec quelques éléments d'information liés à l'utilisation de nos répondants. Pendant nos entretiens, nous avons considéré l'absence de réponse comme une réponse en soi. En tant qu'intervieweur nous avons tâché d'adopter une attitude à la fois sérieuse et empathique. Le répondant devait être en confiance pour se livrer de manière authentique sur son expérience. Si l'informant n'était pas suffisamment exhaustif sur un aspect en particulier, nous avons deux options pour permettre d'approfondir sa réponse.

Nos consignes ou « questions centrales », sont accompagnées de questions de relances. Ces questions visent à réorienter les propos de l'interviewé sur les dimensions de notre recherche, au cas où il ne les aborde pas spontanément. Elles peuvent aussi être une manière de poser le problème ou la question différemment. Dans tous les cas, elles sont une manière de spécifier et/ou de rapprocher le participant de son vécu avec l'assistant. Si les questions de relances issues de notre guide s'avèrent peu pertinentes au vu de

la tournure de la conversation, d'autres questions qui surgissent en cours d'interview peuvent remplir un rôle similaire.

En outre, nous pouvons demander des précisions et/ ou des compléments d'informations sur certains points. Nous prenons cependant le soin de ne pas référer à d'autres interviews que nous avons réalisées. Enfin, nous nous inspirons de l'expérience de Åkerlind (2005) en nous appuyant sur les exemples concrets du phénomène fournis par les interviewés. Même si nous ne sommes pas intéressés par le détail de ces exemples, nous pouvons nous en servir comme médium pour explorer l'expérience, et ces exemples témoignent des aspects du phénomène dont les informants se révèlent conscients.

Enfin, nous avons opté pour des entretiens en ligne, par visioconférence. Cette option présente à nos yeux au moins trois avantages. Le premier réside dans le confort que ça représente pour nos participants, qui gagnent du temps et qui ne doivent pas se soucier d'un quelconque déplacement. Le second relève du caractère rassurant d'un tête à tête à domicile, isolé des perturbations extérieures. Enfin, l'entrevue en ligne permet au chercheur d'enregistrer et de stocker facilement les entretiens pour la phase de retranscription.

2. Recrutement et profils des participants

Pour répondre à la nécessité d'adopter une démarche compréhensive, nous avons décidé de réaliser des entretiens avec un échantillon d'individus interagissant de manière plus ou moins fréquente avec des AV. Afin de maximiser nos chances d'observer des niveaux de conception différents nous avons cherché à varier les profils. Contrairement à beaucoup d'études phénoménographiques, nous ne faisons pas le choix de concentrer nos observations sur un public particulier. Nous faisons l'hypothèse méthodologique, appuyée par d'autres études antérieures (Burbach et al., 2019 ; Thakur & Han, 2018) que la diversification de facteurs tels que l'âge, la formation, le genre, l'assistant utilisé ou encore la fréquence d'utilisation

influence l'expérience d'interaction, et engendre des niveaux de compréhensions variés.

Pour composer notre échantillon nous avons opté pour un questionnaire de recrutement permettant de situer nos participants par rapport à quelques variables liées à leur utilisation des assistants virtuels (cfr. annexe 1). Ce questionnaire nous a par ailleurs permis d'introduire les potentiels participants à notre thématique et à obtenir leur consentement pour l'enregistrement d'une éventuelle interview. Nous avons soumis notre « Google Form » majoritairement sur les réseaux sociaux. Les premières personnes sélectionnées, issues de notre entourage, nous ont ensuite orientés vers d'autres personnes susceptibles de participer à notre recherche. Sur base des 16 réponses obtenues via ce questionnaire, nous avons sélectionné 10 personnes aux profils singulièrement distincts. pour l'interview semi-dirigée. Chacune de ces personnes participe à un entretien. Un récapitulatif des différents profils est repris dans le tableau ci-dessous :

N°	Nom	Sexe	Age	Secteur d'activité	Ancienneté d'utilisation	Fréquence d'utilisation ⁶	AV utilisé
1	David	M	26	Marketing	Entre 3 et 5 ans	Peu fréquent	Google
2	Alix	F	23	Communication digitale	Entre 1 et 3 ans	Fréquent	Siri
3	Isabelle	F	55	Bio-ingénieur	Entre 1 et 3 ans	Fréquent	Google
4	Sacha	M	27	Évènementiel	Entre 3 et 5 ans	Peu fréquent	Siri
5	Sébastien	M	22	Études en littérature	Entre 3 et 5 ans	Peu fréquent	Google
6	Nathan	M	34	Évènementiel	Entre 3 et 5 ans	Très fréquent	Siri

⁶ La fréquence d'utilisation a été approximée en fonction de la fréquence d'utilisation exprimée par les participants par rapport aux recherches "textuelles" (cfr. questionnaire de recrutement)

7	Célie	F	20	Science Politique	Entre 1 et 3 ans	Fréquent	Siri
8	Maxime	M	25	Audiovisu el/ musique	Moins d'un an	Très peu fréquent	Google
9	Paul	M	59	Logistique de transport	Entre 1 et 3 ans	Fréquent	Siri
10	Juliette	F	23	Linguistiq ue	Entre 3 et 5 ans	Peu fréquent	Google

Tableau 1 : profils des participants

Analyse des données

L'analyse des données comporte plusieurs phases. Nous décrivons une à une les étapes du processus d'identification des catégories de description. Nous faisons le pari que les catégories émergeront des données récoltées sur le terrain. Notons néanmoins que les chercheurs ont des opinions diverses sur l'origine des catégories de description. En choisissant d'identifier les catégories pendant la phase de collecte des données et de transcription des entretiens, nous adoptons une approche semblable à celle de Limberg (2008). Pour Marton (1986), ce serait plutôt les "énoncés" des entretiens qui permettraient de révéler les catégories de signification du phénomène étudié dans un certain contexte.

Nous faisons le choix de structurer notre entrevue sans chercher à construire préalablement d'hypothétiques catégories, à partir d'un cadre théorique. Pour assurer la fiabilité et la validité de ce biais méthodologique, et plus généralement de notre enquête, nous détaillons les étapes interprétatives de notre processus, en présentant des exemples pour illustrer ces étapes. Toujours suivant la logique du contrôles de la validité, la recherche phénoménographique exige de la part du phénoménographe qu'il soit capable de défendre les résultats qu'il obtient en les confrontant à ceux d'autres chercheurs. Comme nous travaillons seul et qu'il nous paraît difficile de rentrer directement en contact avec une communauté de chercheurs, nous

confronterons nos résultats à des recherches connexes dans la section de discussion.

De manière plus pragmatique, il est nécessaire de présenter dans quelle mesure les résultats peuvent être considérés comme utiles pour un public visé. Pour ce faire, nous justifierons comment ce travail peut faire office de cadre d'orientation pour des recherches supplémentaires ou pour le développement de démarches de formation ou d'éducation.

Concernant l'analyse de données, nous avons réalisé les entretiens et les avons transcrits. La première phase du processus d'analyse comparative consiste à la lecture et à relecture attentive de l'ensemble des entretiens (Åkerlind, 2005 ; Bowden & Green, 2005 ; Marton and Booth, 1997). Nous avons considéré chaque transcription dans le contexte de l'ensemble des entretiens. Chaque transcription doit en effet prendre son sens en relation avec les autres. Cependant, pour simplifier notre travail, nous avons divisé les transcriptions en plusieurs parties avant d'en chercher ensuite les associations. Les transcriptions ont donc été relues une première fois et les sections qui semblaient pertinentes pour le phénomène ont été surlignées. Nous avons ensuite procédé à des relectures pour nous assurer de n'oublier aucun passage significatif. Les transcriptions et les sections marquées ont été mises en évidence à l'aide de codes, en marge du document. Ceci a été fait afin que les sections puissent être considérées dans le contexte de leurs entretiens d'origine (contexte référentiel) et dans le contexte des autres (contexte structurel). Nous avons ensuite transféré les codes dans un document d'ensemble afin de les comparer. Sur cette base, nous avons identifié plusieurs catégories potentielles. Des codes couleurs ont alors été utilisés pour surligner les passages s'apparentant à des conceptions similaires. Certaines catégories en puissance n'ont pas été retenues, car jugées trop peu pertinente au regard de l'ensemble des entretiens.

1. Développement de catégories

L'analyse des données que nous avons décrite a permis de dégager cinq manières qualitativement différentes dont les utilisateurs ont vécu leur interaction avec leur AV. Nous appellerons ces différentes façons de comprendre le phénomène, les « catégories de descriptions ». Elles sont donc directement issues de leur témoignage et leur expérience globale, plutôt que des réponses à des questions spécifiques.

Les 5 catégories de descriptions sont énumérées ci-dessous. Nous les détaillons ensuite en les illustrant à l'aide de citations d'utilisateurs. Dans un deuxième temps, les relations entre les catégories sont présentées dans un tableau récapitulatif qui les confronte selon 4 paramètres distincts (tableau 2).

- A. L'interaction comme un moyen de se divertir.
- B. L'interaction vocale comme une alternative éventuelle à l'interaction gestuelle.
- C. L'interaction comme processus routinier pour des tâches spécifiques et récurrentes.
- D. L'interaction comme un moyen permettant une pratique multitâche.
- E. L'interaction comme un moyen d'organisation et de gestion de la vie privée et professionnelle.

1.1 Catégories A : Un moyen de divertissement

Dans la catégorie A, l'interaction avec l'assistant vocal est considérée comme un moyen de divertissement. L'utilisateur est motivé par la découverte et l'exploration des moyens mis en place par les assistants virtuel pour simuler la conversation. L'assistant vocal est considéré comme un jeu qui laisse libre cours à des expérimentations. L'utilisateur explore, il sort des sentiers battus. Il ne cherche pas à commander à partir de phrases prédéfinies, il préférera au contraire laisser parler son imagination pour tester la « répartie » de l'assistant virtuel. En ce sens, l'assistant sera régulièrement considéré comme un gadget plutôt que comme un outil. L'irrégularité perçue

dans la qualité des réponses fournies semble entraver l'idée d'une utilisation plus « sérieuse ». L'interaction avec l'assistant n'est pas considérée comme permettant la résolution de tâches ou est perçu comme inutile à cet effet.

« (...) c'est marrant d'avoir un aspect un peu humain derrière. Parfois tu vas être surpris parce qu'il va faire une blague ou il va avoir une bonne répartie mais juste après il va te dire "je n'ai pas compris la question" alors que ce n'était même pas une question que tu faisais... » (Maxime)

« Je ne lui demande pas de compléter des tâches dans ma vie... Absolument pas. (...) J'ai l'impression qu'il se limite à juste fonctionner comme un moteur de recherche. (...) Pour moi ça ne coûte rien et c'est limite plus rapide et plus pratique de taper sur l'écran que d'attendre et d'interagir avec l'assistant vocal. » (Sébastien)

Dans cette catégorie, on pourrait apparenter l'utilisateur à « l'explorateur » dans la typologie des joueurs de Richard Bartle (1996). L'utilisateur peut aussi bien vouloir tenter d'explorer l'étendue du système que la profondeur de ce dernier. Dans le premier cas, cela pourrait par exemple se manifester par une multiplication de questions visant à cartographier les capacités de réponses :

«Encore aujourd'hui je l'utilise beaucoup mais pour me marrer (...) Je lui pose des questions absurdes, j'essaye de lui faire dire des trucs sur lui-même, je lui demande de faire des blagues. » (Sébastien)

Dans le deuxième cas, l'utilisateur pourrait par exemple vouloir tenter de converser avec l'assistant sur un sujet précis ou tenter d'échanger des opinions. Dans les deux cas, cette conception pousse à s'interroger sur les caractéristiques anthropomorphiques du système. Deux utilisateurs ont également évoqués leur curiosité concernant la conversation inter-assistants. Au-delà du divertissement que cela constitue, cela participe de cette même volonté d'exploration des capacités du système. Dans cette catégorie, nous remarquons également que l'interaction est souvent pensée comme pouvant inclure un troisième sujet. Cela peut être un groupe de personnes ou une personne seule passive ou active dans l'interaction. Si le troisième sujet est

plutôt actif dans l'interaction, celle-ci prendra alors son sens à travers la contribution de chacun. Le couple « utilisateur-assistant » peut également constituer une forme de démonstration dont le troisième sujet est témoin (passif).

*« Ça peut faire rire en société. Genre on a passé une soirée comme ça avec des potes chez moi à parler à Google et j'ai demandé des trucs, c'était assez marrant, **on** jouait avec lui. » (Sacha)*

*« Moi je ne l'utilise pratiquement jamais tout seul, ça sera toujours en compagnie de quelqu'un, genre mes potes ou avec une personne plus âgée comme mon père **pour lui montrer un peu** : "regarde ce que je fais avec mon téléphone", tu vois ? » (Sébastien)*

On remarque que cette conception pousse l'utilisateur à adopter un ton plutôt familier et des structures de phrases naturelles. L'utilisateur ne cherche pas à adapter son discours au système. Au contraire, voir à quel point le système peut s'adapter au discours de l'utilisateur fait partie intégrante du processus de divertissement.

« Donc, soit je lui parle de façon pratique pour obtenir une réponse directe, soit je lui parle de tout et de rien pour rigoler avec des phrases plus complètes. » (Sébastien)

Enfin, cette première catégorie de description entraîne également un affaiblissement progressif de l'utilisation au fil du temps. L'interaction est perçue comme de moins en moins divertissante. Une fois que le « jeu » est considéré comme étant exploré, il ne reste plus beaucoup de raisons de continuer à interagir avec l'assistant vocal. De la même manière, l'absence de « nécessité » est avancée comme un facteur de désintérêt progressif.

« (...) c'était plus du jeu qu'autre chose, tu vois ? Voir comment ça fonctionne, tu rigoles et après c'est vrai que tu l'utilises assez peu. (...) par définition jouer ce n'est pas quelque chose de nécessaire tu vois ? (...) je l'ai utilisé de moins en moins en fait. Au fur et à mesure que la curiosité s'est estompée j'ai arrêté de l'utiliser. C'est étonnant

parce que généralement quand tu as une technologie à ta disposition ton utilisation s'améliore au fil du temps. » (Maxime)

« Je définirais ça plutôt comme un jouet (...) Genre si on devait supprimer cette fonctionnalité de mon téléphone demain je serais un peu déçu parce que je l'utilise parfois mais je pourrais clairement m'en passer et peut être que j'éviterais même de perdre du temps. » (Sébastien)

1.2 Catégorie B : Une alternative à l'interaction gestuelle

Dans la catégorie B, l'interaction avec l'assistant vocale est considéré comme une alternative éventuelle à l'interaction gestuelle. Bien que l'interaction gestuelle reste privilégiée dans la majorité des cas, l'utilisateur conçoit l'interaction vocale comme pouvant être « avantageuse » pour l'exécution de certaines tâches simples et occasionnelles.

« je cherche des informations qui sont quand même faciles d'accès mais qui prendrait plus de temps à chercher en les tapant. » (Juliette)

L'utilisateur a un rôle particulièrement actif dans la sélection des tâches qu'il effectue avec l'un ou l'autre type d'interaction. Le recours à l'assistant vocal pour effectuer une tâche particulière n'est pas systématique. Il dépend notamment des circonstances et de la connaissance des capacités du système.

« Pour moi ça va rester des phrases assez basiques, des choses où je me dis "là, il doit y avoir une manière de le dire." » (Sacha)

« on ne sait pas toujours quel mot employer, ou plutôt on ne sait pas toujours ce à quoi la machine pourra répondre... Pour ça, on est obligé de demander : "Google, qu'est-ce que tu peux faire ?" » (Paul)

« le truc que je fais le plus c'est de demander de changer de musique. (...) quand j'veux pas me déplacer et que j'suis tranquille dans ton lit je peux utiliser l'assistant quoi. » (David)

De manière générale, l'alternative vocale semble être considérée intéressante lorsqu'elle permet un gain de temps. La facilité d'usage ou « l'efficacité » sont mentionnées dans une moindre mesure. L'utilisateur se considère capable d'évaluer les tâches pour lesquelles il va gagner du temps avec l'assistant vocal. Il cherche à optimiser son utilisation.

« Si je me rends compte que je peux faire quelque chose plus vite par la voix je le fais par la voix, si je peux le faire plus vite par moi-même, ben je le ferai avec mes mains. » (David)

« J'utilise principalement l'assistant vocal pour gagner du temps. (...) il y a certaines tâches pour lesquelles je vais commencer à lui faire confiance. Alors que pour d'autres tâche je vais peut-être préférer prendre mon GSM et faire moi-même sur l'écran. (...) Par exemple si je veux faire une recherche Google c'est rare que j'utilise l'assistant parce que je sais qu'il y a une chance sur deux qu'il ne comprenne pas ma requête ou bien qu'il ne trouve pas la réponse. » (Isabelle)

Si cette catégorie considère que l'interaction vocale peut parfois consister en un gain de temps significatif, ce n'est pas pour autant que les pratiques sont définitivement figées et systématiques. Parfois l'utilisateur passera d'une modalité à l'autre. Toujours dans une logique d'optimisation, la modalité tactile semble rester le moyen le plus évident de corriger les incompréhensions qui résulte de l'interaction vocale.

« faire des corrections c'est vraiment compliqué avec la voix tu vois ? Du coup je préfère m'arrêter et remplir l'adresse à la main. Je n'ai pas encore trouvé le bon moyen pour dicter mes itinéraires. » (Sacha).

« Quand ça ne marche pas j'utilise mes doigts. Je ne cherche pas à comprendre pourquoi ça ne marche pas. » (Alix)

Certains participants ont également rappelé l'importance de se « souvenir » de la présence de l'assistant vocal. Son utilisation semble donc aussi conditionnée par le fait d'y penser, ce qui n'est manifestement pas si évident pour tous les utilisateurs.

« Moi je ne l'utilise pas énormément aussi parce que j'oublie qu'il est là en fait. » (Célie)

« je vais l'utiliser beaucoup pendant une petite période de temps, genre le temps d'une journée, et puis plus du tout. Genre pendant un moment ça va être frais dans ma tête donc je vais y penser et je vais l'utiliser (...) Mais après ça arrive que je ne l'utilise plus du tout pendant une semaine quoi. Parce que finalement la normalité ça reste l'utilisation classique, l'utilisation tactile quoi. » (David)

Enfin, cette catégorie sous-tend des interactions assez irrégulières, aussi bien en termes de récurrences qu'en termes de formulation. L'interaction reste en effet relativement occasionnelle, et de ce fait, elle peut générer davantage d'incompréhensions. Bien que l'utilisateur sache généralement quand il peut gagner du temps avec l'assistant, il ne sait pas toujours pour autant pourquoi et comment s'adresser à lui. L'utilisateur reconnaît volontiers les performances de la technologie mais il regrette le manque de visibilité sur son utilisation concrète.

« c'est très rare qu'il [l'assistant vocal] ne comprenne pas. Même quand il y a du bruit ou quoi il comprend. Maintenant ce qui m'embête plus c'est quand il ne comprend pas et qu'on ne sait pas pourquoi. Du coup on ne sait pas si on doit bêtement répéter la phrase ou bien si on doit la reformuler. » (David)

1.3 Catégorie C : Un processus routinier

Dans la catégorie C, l'interaction est perçue comme un outil du quotidien qui permet d'effectuer de petites tâches spécifiques et récurrentes. Cela peut se réaliser sous forme de contrôles de routine ou de tâches simples et périodiques. Les utilisateurs manifestent ici la volonté de recourir à l'assistant vocal de manière plus systématique à des moments clés de la journée.

« c'est vrai que depuis je l'utilise souvent, ça marche bien. Je parle à mon téléphone ! Le matin je lui demande de consulter la météo et les infos. » (Isabelle)

« Moi je le conçois un peu comme un outil du quotidien qui permet de permet de te faire exécuter des tâches à ton téléphone. » (Sacha)

« Pour le reste j'ai commencé à l'adopter aussi pour d'autres petites choses du quotidien, comme pour demander la météo ou le programme télé par exemple... ou pour demander des recettes, des petites choses comme ça quoi. » (Paul)

L'interaction avec l'assistant vocal est considérée progressivement comme privilégié par rapport à l'interaction gestuelle pour un certain nombre de petites tâches très limitées. L'utilisation de l'assistant se limite généralement à ces usages. L'utilisateur connaît les tâches pour lesquelles il mobilise l'assistant et il ne cherche pas à dépasser les capacités du système. Il ne cherche a priori pas non plus à actualiser ses usages avec le temps.

« (...) moi je l'utilise pour les choses pour lesquelles je suis sûr que ça marche, tu vois ? (...) j'ai toujours utilisé les assistants pour les mêmes trucs et franchement il n'y en a pas beaucoup quand j'y pense. » (Sacha)

« (...) pour l'utiliser correctement de base, je posais seulement les phrases type que Siri proposait. Et du coup j'ai continué à poser ces questions-là pour mettre un réveil, mettre un rappel ou passer un appel, ou écrire un message ou quoi. (...) depuis que j'ai vu pour quoi et comment je pouvais l'utiliser et bien je n'ai pas cherché à lui faire faire d'autres choses. En plus, les questions je trouve qu'il les comprend pas du tout du coup je préfère lui faire faire des commandes quoi. » (Alix)

« (...) je suppose qu'avec le temps la liste de choses [de commandes possibles] s'agrandit mais moi à priori je vais continuer à l'utiliser pour les choses que j'ai l'habitude de faire. Je ne vais pas aller me renseigner tous les mois pour savoir si l'application a évolué quoi. » (Paul)

Cette catégorie conçoit l'interaction avant tout comme « facilitatrice » de petites tâches. Contrairement à la catégorie précédente qui donnait une priorité au gain de temps, celle-ci aura plus tendance à considérer avant tout l'aspect pratique de l'interaction. L'usage peut être considéré comme facilité

par le fait d'éviter l'interface visuelle ou par le fait d'éviter le mouvement de saisie du téléphone.

« Mais moi personnellement j'utilise l'assistant aussi quand j'ai les mains libres. Dans tous les cas je trouve que c'est assez confortable l'idée de ne pas avoir de menus, de boutons et tout ça. » (Paul)

Q : Ça te permet de gagner du temps ?

« Je ne trouve pas non, ce n'est pas un gain de temps.(...) je dis : "Hey Siri" et puis je commence à parler et c'est vraiment ça que je trouve pratique. S'il faut sortir son téléphone et appuyer sur un bouton avant de parler de vois pas trop l'intérêt. » (Célie)

Dans cette catégorie, on observe moins de variabilité dans le type d'interaction mais plus de récurrences dans la formulation des commandes. Les utilisateurs ont une idée bien claire de la manière d'exprimer leurs commandes car ils les réitèrent régulièrement. Ces commandes sont d'ailleurs parfois directement issues des suggestions des assistants vocaux, ce qui ne laisse a priori aucune place aux incompréhensions (Ex : « Quelles sont les conditions de circulation ? », « Mets les infos. », « Quel temps fait-il aujourd'hui »⁷, etc.) .

1.4 Catégorie D : Une pratique multitâche

Cette catégorie exprime une conception de l'interaction avec l'assistant vocal comme un moyen permettant une pratique multitâche. Elle regroupe des caractéristiques des deux catégories précédentes et en présente d'autres. Similairement à la catégorie C, l'interaction est perçue avant tout comme pratique, voir même essentielle dans certains cas. Elle trouve son intérêt dans des conditions particulières, pour lesquelles l'utilisateur est déjà engagé dans une activité qui mobilise ses mains. Les usagers considèrent ici que l'interaction leur permet de dédoubler leur activité. Cela peut répondre à un besoin, mais ça peut aussi représenter une forme de sécurité, ou de dépannage.

⁷ Exemples issus des suggestions de *Google Assistant* (Date : 05/12/2021)

« [je l'utilise] en voiture aussi. Je ne peux pas prendre mon téléphone en main et donc voilà, si j'ai envie d'appeler quelqu'un je peux le faire sans devenir un danger pour les autres automobilistes. C'est aussi une forme de sécurité en fait. » (Isabelle)

Q : Y-a-t-il des circonstances d'utilisation plus adaptées ?

« Je dirais dans la rue quand j'ai les AirPods ou bien quand je cours, quand je fais du sport. Dans des situations où ne t'as pas envie de t'arrêter ou t'as envie de faire plusieurs choses à la fois. » (Célie)

« En bien ce sera particulièrement utile quand je cuisine par exemple... En fait il y a des situations où tu n'as pas vraiment le choix. Quand tes mains sont mouillées ou quand tu portes des gants par exemple, ce n'est pas possible d'utiliser le tactile du téléphone.(...) ça vient compléter un manque d'une certaine manière. » (Paul)

« Moi c'est souvent quand je suis occupé aussi que je vais l'utiliser. C'est plus pour dépanner qu'autre chose. C'est un outil de dépannage aussi. » (Alix)

Cette conception est également caractérisée par une grande confiance à l'égard de l'assistant. Une situation où les mains ne sont pas disponibles implique un contrôle total par la voix. Dans ce contexte, un utilisateur mentionne l'importance de l'interaction à deux tours :

« en termes de conversation ce sera surtout dans le cas de "l'Apple Car⁸" par exemple ou Siri va lancer une conversation du style : "Telle personne appelle, est-ce que vous voulez répondre ?" Du coup il inverse un peu le processus et c'est lui qui s'adresse à moi dans ce cas-là.» (Nathan)

Ces interactions à deux tours (ou plus) peuvent également présenter un obstacle dans la pratique multitâche. Elles sont en effet compliquées à générer par l'unique usage de la voix. Quand une incompréhension survient, une suspension de l'activité est alors parfois nécessaire.

⁸ CarPlay est une version embarquée d'iOS pour les véhicules. Elle permet d'utiliser Siri directement sur le tableau de bord de la voiture.

« [à propos de ses commandes] faire des corrections c'est vraiment compliqué avec la voix, tu vois ? Du coup je préfère m'arrêter et remplir l'adresse à la main. Je n'ai pas encore trouvé le bon moyen pour dicter mes itinéraires. » (Sacha)

Par ailleurs, la pratique multitâche telle qu'elle est pensée par les utilisateurs se conjugue volontiers avec d'autres supports. Certains utilisateurs ont laissé entendre que des accessoires intermédiaires pouvaient venir renforcer ou faciliter leur interaction avec l'assistant vocal.

« (...) j'utilise l'assistant de mon téléphone mais **indirectement**. Genre je l'utilise dans la voiture, et dans la voiture c'est connecté à mon tel quoi (...) Je ne fais pas le con à prendre le tel au volant et tout ça (...) j'appuie sur le bouton de l'assistant vocal sur mon volant et je lui demande de mettre un son ou un itinéraire. » (Sacha)

« En fait depuis que j'ai les **AirPods** je l'utilise plus parce que ça me permet de commander la musique à distance par exemple, et il y a un micro sur les AirPods. Du coup je peux par exemple régler le son facilement sans devoir sortir le téléphone. » (Célie)

1.5 Catégorie E : Un moyen d'organisation et de gestion

L'interaction avec l'assistant vocal peut être vu comme un moyen d'organisation et de gestion de la vie privée et professionnelle. Cette conception se rapproche de celle d'un assistant dans son sens littérale. Selon le Larousse en ligne, l'assistant peut être défini comme suit : « personne qui en seconde une autre dans un travail⁹ ». Cette idée « d'accompagnement » dans l'activité est plus prégnante dans cette catégorie que dans les précédentes. L'utilisateur démontre une volonté d'interagir avec l'assistant pour déléguer une partie de sa charge de travail. En plus du déchargement mental, l'utilisateur peut vouloir interagir pour étendre ses capacités notamment mémorielles.

⁹ Larousse. (2021). assistant, assistante. Dans Dictionnaire en ligne. Consulté le 6 décembre 2021 sur <https://www.larousse.fr/dictionnaires/francais/assistant/5847>

« [je l'utilise] quasi exclusivement dans le cadre professionnel. J'utilise vraiment l'assistant comme un aide-mémoire. Et du coup quand je suis en mouvement, quand je suis en déplacement, dès que j'ai une info qui me passe en tête, dès que j'ai une info dont je dois me souvenir, j'utilise énormément Siri pour me mettre des rappels et des missions à faire plus tard.(...)ça me permet de me décharger mentalement de choses très rapidement. C'est de la charge mentale en moins pour moi. » (Nathan)

« Mon père est assez geek (...) on a de plus en plus de produits connectés à la maison (...) on a "Google Assistant" configuré sur le Google home (...) on doit demander à Google d'allumer l'eau chaude. Comme ça on peut faire des économies d'eau chaude parce qu'on ne doit pas chauffer en permanence. On peut aussi mettre certaines heures pour qu'il le fasse. (...)Du coup il faut demander à Google si j'ai envie de prendre ma douche à 14h, et le fait de devoir verbaliser ça nous permet de ne pas oublier. Enfin on le fait plus facilement, juste avec une petite phrase, du coup c'est moins contraignant et c'est plus naturel. » (Sacha)

Cette conception de l'interaction bouleverse plus profondément les habitudes de l'utilisateur. Elle peut être à l'origine de nouvelles manières d'appréhender la gestion du quotidien. L'interaction permet d'élargir le champ de conscience organisationnel.

« Je le fais pour tout [utiliser l'assistant]. Ça peut être genre "appeler Jean-Pierre", ou fixer un rendez-vous, j'ai pris l'habitude de demander à Siri de le faire. Et avant que j'utilise Siri je n'utilisais même pas l'agenda électronique. C'est vraiment avec Siri que j'en ai vu l'utilité.(...) je ne sais pas comment formuler ça mais il permet de pouvoir tout mettre sous forme de rappel sans devoir euh... de manière dématérialisée » (Nathan)

Cette catégorie décrit une conception qui semble déjà s'ancrer dans le réel pour certains, alors que pour d'autres cela relève encore du fantasme. Cette conception ne peut pas être la leur dans l'état actuel des choses, car leur utilisation concrète ne leur a pas permis un tel abandon d'eux-mêmes dans cette technologie. Plusieurs facteurs, dont la plupart ont été mentionnés dans la première partie de ce mémoire, semblent encore entraver l'idéal qu'ils décrivent.

« [à propos de l'assistant vocal du futur] j'aimerais bien avoir une sorte d'assistance tu vois ? Un assistant lié au boulot alors plutôt... Une sorte de personne qui m'aide pour m'organiser dans le cadre de mon travail tu vois ? (...) Je préférerais mille fois pouvoir converser un assistant compétant, genre qui aurait en mémoire ce que moi j'ai oublié ou que je ne sais pas. » (Célie)

« [à propos de l'assistant vocal du futur] ça devrait être un peu comme dans Iron Man tu vois ?(...) genre littéralement le truc il est partout et tu lui demandes un truc et il donne l'info.(...) Tu vois genre t'es dans ta maison où que tu sois, il peut t'entendre et "OK ça va Sacha ? Tu veux quoi ?" » (Sacha)

Enfin, cette catégorie se caractérise par une haute estime du système et de ses facultés. L'interaction devient presque indispensable et systématique aux yeux de l'utilisateur. L'utilisateur interagit avec confiance et il a une bonne maîtrise de l'outil qui l'accompagne.

« Siri me comprend bien dans l'ensemble. Je trouve que c'est une technologie qui marche super bien et je suis étonné qu'il n'y ait pas plus de gens qui l'utilise. » (Nathan)

2. Espace de résultat

2.1 Relations entre les catégories

Nous proposons le tableau récapitulatif suivant qui compare les catégories de descriptions selon l'objectif principal, le champ d'interaction, le rôle de l'utilisateur et le style d'interaction principale :

Cat.	A	B	C	D	E
Éléments clés	L'interaction comme un moyen de se divertir	L'interaction vocale comme une alternative éventuelle à l'interaction gestuelle	L'interaction comme processus routinier	L'interaction comme un moyen permettant une pratique multitâche	L'interaction comme un moyen d'organisation et de gestion

Objectif principal	<ul style="list-style-type: none"> - Tester la “répartie” de l’assistant. - Entretenir la distraction. 	<ul style="list-style-type: none"> - Considère avant tout le gain de temps dans l’interaction - Interaction alternative 	<ul style="list-style-type: none"> - Considère avant tout la facilité dans l’interaction - Interaction complémentaire 	<ul style="list-style-type: none"> - Volonté de dédoublement de l’activité - Sécurité ou dépannage - Actions ciblées 	<ul style="list-style-type: none"> - “Assistance” dans sa signification littérale - Déchargement mental/ allégement du travail cognitif
Champ de l’interaction	<ul style="list-style-type: none"> - Pas ou peu d’exécution de tâches. - Tentative de conversation. - Jusqu’où l’anthropomorphisme est-il poussé ? 	<ul style="list-style-type: none"> - Interaction gestuelle reste privilégiée - Interaction circonscrite par la connaissance des capacités du système¹⁰ - Petites tâches diverses et variés 	<ul style="list-style-type: none"> - Interaction bien délimitée - Nombre limité de tâches/ contrôles de routine 	<ul style="list-style-type: none"> - Interaction limitée aux contextes impliquant les mains 	<ul style="list-style-type: none"> - Interaction comme réflexe, caractère omniprésent - Accompagne l’usager dans son quotidien
Rôle de l’utilisateur	<ul style="list-style-type: none"> - L’utilisateur comme « explorateurs » de la technologie. - Volonté de se faire “surprendre” par l’assistant. - Utilisateur tributaire des capacités du système¹¹ - Duo assistant-utilisateur comme acteur social 	<ul style="list-style-type: none"> - L’utilisateur comme “optimisateur” de l’interaction - Rôle actif de l’utilisateur dans la sélection de tâches à réaliser - L’utilisateur doit se rappeler de la présence de l’assistant 	<ul style="list-style-type: none"> - L’utilisateur « coutumier » dans ses usages - Connait les tâches pour lesquels mobilise l’assistant - Ne cherche pas à dépasser les capacités du système qu’il conçoit 	<ul style="list-style-type: none"> - L’utilisateur est déjà engagé dans une activité - Confiance importante à l’égard de l’assistant 	<ul style="list-style-type: none"> - L’utilisateur étend ses capacités grâce à son assistant vocal - L’organisation personnelle de l’utilisateur dépend de l’assistant
Style d’interaction principal	<ul style="list-style-type: none"> - Très occasionnelle - Interaction souvent en présence de témoins - Contexte amical ou familial - Primauté à l’innovation dans les questions 	<ul style="list-style-type: none"> - Occasionnelle - Formulations parfois incertaines - Davantage d’incompréhensions - Obstacles et réticences 	<ul style="list-style-type: none"> - Récurrente - Formulation répétitive des commandes - Interaction qui s’opère à des moments précis de la journée - Contexte informel 	<ul style="list-style-type: none"> - Peut combler un manque - Contrôle total du téléphone par la voix (y compris interactions à plusieurs tours) 	<ul style="list-style-type: none"> - Presque indispensable et systématique - Très fréquente - Modifie les habitudes - Phrases complètes, plusieurs informations

Tableau 2 : Relations entre les catégories

¹⁰ Le nombre de tâches entreprises dépend du nombre de tâches que l’utilisateur conçoit comme potentiellement réalisables et utiles avec l’AV.

¹¹ Le divertissement de l’utilisateur sera conditionné par les réponses que le système est capable de fournir

Pour la saisir dans son ensemble, la notion de conception nécessite encore quelques précisions. Larsson (1986) l'a décrite comme suit :

“What (to the individual) is implicit and doesn't need to or can be said as it has never needed to be reflected over and serves as the foundation for our reasoning.”(Larsson, 1986)

Cette définition vient compléter celle de Marton & Pong (2005) selon laquelle la conception est avant tout la manière avec laquelle les gens comprennent un phénomène particulier ou un aspect du monde qui les entoure. Comme nous l'avons brièvement évoqué précédemment, la notion de conception comporte deux aspects :

- L'aspect « référentiel ». Il renvoie à la signification de la conception. Cette dimension se concentre sur ce que les individus expérimentent (le « quoi »).
- L'aspect « structurel ». Il montre la combinaison spécifique des caractéristiques qui ont été prises en compte. Cette dimension se focalise sur la façon dont les individus pensent au phénomène (le « comment »).

Le tableau suivant permet de faire le distinguo entre les deux aspects à partir du cas concrets des conceptions observées dans le cadre de notre recherche :

Conception	Aspect référentiel	Aspect structurel	Horizon
A	L'interaction comme un moyen de <i>divertissement</i>	Centré sur les <i>artifices</i> du système permettant une pratique ludique	Externe
B	L'interaction vocale comme une <i>alternative</i> éventuelle à l'interaction gestuelle	Centré sur les <i>fonctions</i> du système permettant d'optimiser l'interaction	Externe

C	L'interaction comme <i>processus routinier</i> pour des tâches spécifiques et récurrentes	Centrée sur les <i>habitudes</i> dans les <i>usages</i> de l'utilisateur	Interne/externe
D	L'interaction comme un moyen permettant une pratique <i>multitâche</i>	Centrée sur <i>l'activité</i> de l'utilisateur et les <i>circonstances</i> d'utilisation	Externe
E	L'interaction comme un moyen <i>d'organisation</i> et de <i>gestion</i> de la vie privée et professionnelle	Centrée sur <i>l'approche</i> adoptée par l'utilisateur pour faire face aux problèmes du quotidien	Interne

Tableau 3 : Aspect référentiel/structurel

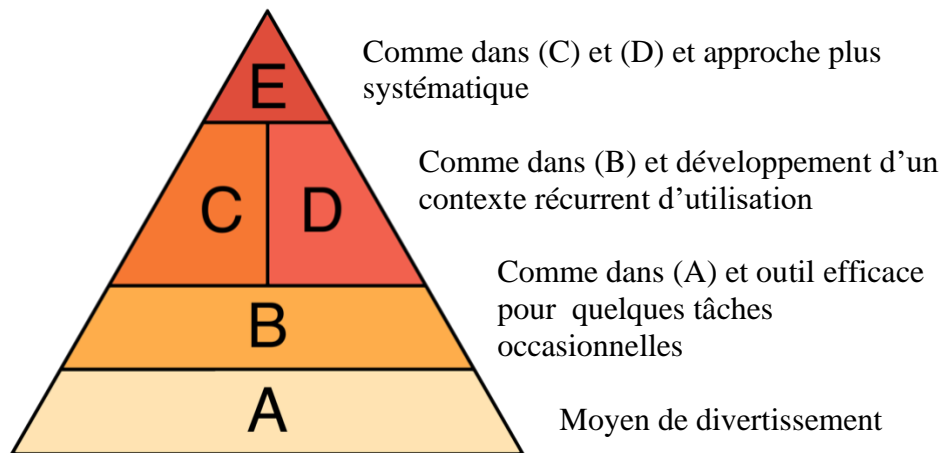
Le tableau est complété par les dimension interne et externe des différentes conceptions. Barnard, et al. (1999) expliquent les termes d'horizon interne et externe comme suit :

“Internal horizon refers to how component parts of the phenomenon are understood and are related to each other. External horizon refers to the way in which the phenomenon is delimited from and related to its context. The term delimit refers to the suspension of attributed meaning or definition from a preconditional perspective.”

(Barnard et al., 1999)

Dans le contexte de cette étude, l'horizon interne fait référence à la façon dont les utilisateurs conceptualisent l'interaction avec l'assistant vocal et forme une compréhension cohérente de la façon dont les deux parties (l'utilisateur et l'AV) sont liées. L'horizon externe concerne la façon dont les méthodes d'impact sont liées aux contraintes du contexte et de l'environnement.

2.2 Organisation hiérarchique



Les catégories de description dans les études phénoménographiques peuvent avoir des relations logiques entre elles (Marton & Booth, 1997). Le chercheur ne doit pas seulement mettre le doigt sur différentes « significations », mais aussi rechercher une structure logique qui les relie. Ces relations sont supposées constituer une hiérarchie structurelle qui ordonne les catégories de la plus « simple » à la plus « complexe » (Åkerlind, 2005). Cette terminologie mérite d'être commentée, car elle peut être trompeuse. La catégorie d'ordre supérieur est considérée comme plus complexe car plus complète par nature, c'est-à-dire qu'elle comprend des éléments des catégories d'ordre inférieur.

Dans le cas qui nous occupe, la dernière catégorie (catégorie E) est hiérarchiquement supérieure aux quatre catégories précédentes (catégories A, B, C et D). Elle emprunte en effet les caractéristiques des conceptions antérieures en rajoutant une dimension à celles-ci. L'interaction est considérée non seulement comme un moyen de divertissement à certains égards, mais aussi comme une alternative pertinente à l'exécution de tâches dans plusieurs cas de figure, un processus qui peut devenir récurrent et/ou permettre de faire plusieurs choses à la fois, et enfin comme un moyen plus structurel d'organisation et de gestion. Nous précisons que le niveau de conception (E) n'est pas une fin en soi. On dira simplement qu'il s'agit d'une conception plus « complète » que les autres, du fait qu'elle peut « inclure »

les catégories inférieures. A l'inverse, la catégorie (A) ne laisse pas de place à la conception d'exécution de tâches, et encore moins à celle d'une interaction située et systématique. On dira qu'elle est moins complexe. Dans cette tentative d'organisation hiérarchique, nous avons ordonné les catégories (C) et (D) sur un même niveau de complexité. Nous considérons que la catégorie (D) n'est pas nécessairement d'ordre supérieur. Comme la catégorie (C), elle conçoit l'utilité pour des tâches ciblées et récurrentes et elle priorise l'aspect pratique dans l'interaction. Seule la conception du contexte et de l'objectif d'utilisation diffère entre ces deux catégories, ce qui ne représente pas, à nos yeux, des critères suffisants pour les hiérarchiser. Nous avons néanmoins tenu, par l'usage des couleurs, à marquer une distinction entre les deux catégories. En effet, pour certains utilisateurs, la conception (D) a la particularité de combler un besoin, ce qui pourrait constituer un niveau de complexité supérieur. Cependant, comme cette conception n'implique pas nécessairement cette caractéristique nous avons estimé qu'elle ne devrait pas engendrer un étage supplémentaire, mais qu'elle était suffisante pour être signalée.

3. Identification de thèmes

En plus de l'analyse phénoménographique, les transcriptions ont fait l'objet d'une analyse thématique. Le but était de saisir les thèmes concernant l'utilisation de l'assistant virtuel, au-delà des conceptions, et d'analyser l'usage rapporté. Ces thèmes sortent du cadre de l'analyse phénoménographique, mais ils fournissent tout de même de nouvelles perspectives auxquelles nous pensons devoir accorder de l'importance. Saldaña (2012) affirme que l'analyse thématique convient aux chercheurs qualitatifs débutants, lorsque le chercheur a déjà une idée générale de ce qu'il doit étudier dans les données. Il distingue les « catégories » des « thèmes » de cette manière :

“Think of a category as a word or phrase describing some segment of your data that is explicit, whereas a theme is a phrase or sentence describing more subtle and tacit processes.”(Saldaña, 2012).

Pour faire émerger les thèmes dans nos entretiens, l'analyse thématique à laquelle nous procédons se veut conforme à la théorie de Johnny Saldaña. Selon lui, de la même manière que pour les catégories de descriptions, des codes peuvent aussi être employés pour mettre en avant des segments qui révèlent des processus tacites et récurrents. Ces codes peuvent ensuite être regroupés en catégories qui forment des thèmes, lesquels peuvent à leur tour être utilisés pour construire une théorie. Nous avons donc dupliqué notre document de transcription et nous sommes repartis à zéro. Nous n'avons plus cherché à identifier des catégories de descriptions dans les données mais plutôt des segments thématiques. Cependant, de la même façon que pour la recherche de catégories, les thèmes doivent être saisis en les absorbant dans leur ensemble plutôt qu'être analysés lignes par lignes (Dey, 1993, référencé par Saldaña, 2012). Nous avons donc surligné notre document à l'aide de codes couleurs (marquant les similitudes entre les segments) et dénommer les segments identifiés à l'aide de codes synthétisant l'idée principale. Nous avons procédé à un codage à la fois ciblé et ouvert. Le codage ciblé consiste à préétablir des codes à partir du guide d'entretien, et rechercher ensuite les données qui se rapporte à ceux-ci. Par exemple, en interrogeant nos participants sur les incompréhensions qu'ils rencontrent avec le système, nous avons produits des points de vues divers sur cette thématique particulière. Le codage ouvert consiste quant à lui à créer une liste de codes pendant le codage. Les codes émergent donc de l'analyse des données.

Dans les deux cas, nous avons considéré la valeur des segments identifiés au regard des dimensions la recherche et des questions d'entretiens. Si un segment jugé intéressant ne correspondait pas assez bien à un code couleur existant, un nouveau code couleur était créé. S'ils étaient suffisamment similaires, certains codes ont été quant à eux fusionnés. De cette manière, nous avons pu former plusieurs thèmes provisoires. Une fois bien définis, ces thèmes ont été nommés.

Cinq thèmes centraux ont émergé des transcriptions, comme l'indique le tableau ci-dessous. Ces thèmes s'inscrivent dans le cadre des attentes globales des utilisateurs en matière d'interaction avec l'assistant vocal, ainsi que des facilitateurs et des obstacles qu'ils perçoivent dans son utilisation. Chaque thème est discuté séparément ci-dessous en relation avec les perceptions de l'ensemble des utilisateurs sur ce phénomène.

Thèmes	Codes
Évolution de l'utilisation	<ul style="list-style-type: none"> ▪ Les usages s'actualisent peu ▪ Intensification de l'utilisation avec un assistant performant
Interaction en public	<ul style="list-style-type: none"> ▪ Menace pour l'intimité ▪ Embarras liés au bruit environnant ▪ Peu d'utilisateurs vus en situation d'interaction
À propos du sentiment de contrôle	<ul style="list-style-type: none"> ▪ Influence sur l'utilisateur ▪ Réponses fondées sur le jugement, fatalité ou finalité ? ▪ Interrogations sur le déclenchement de l'assistant
À propos de la conversation utilisateur - assistant	<ul style="list-style-type: none"> ▪ Recherche de fluidité ▪ Réticences liées à l'humanisation du système
À propos de l'in(compréhension)	<ul style="list-style-type: none"> ▪ Discernement entre transcription et commandes ▪ Origines diverses des incompréhensions ▪ Réactions face aux incompréhensions

Tableau 4 : Thématiques dans la perception de l'interaction par le groupe d'intérêt

3.1 Évolution de l'utilisation

La perception des utilisateurs concernant leur utilisation a ceci en commun qu'elle est globalement perçue comme peu évolutive. Nos participants considèrent qu'il ont rapidement adopté l'usage qui est le leur, le temps influençant très peu. Un utilisateur met le doigt sur un aspect

important ; contrairement à d'autres technologies dont il a l'usage, il n'a pas développé pour l'AV la volonté, voir le besoin, « d'élargir » ses habitudes :

« Mon utilisation ces dernières années n'a pas du tout évolué (...) c'est bizarre parce que j'utilise beaucoup d'objets connectés pour lesquels j'ai l'impression que mon utilisation change au fil du temps mais les assistants vocaux par vraiment. » (Sacha)

Cette observation coïncide avec celle de Budiu (2019) : une fois que les utilisateurs ont décidé que l'assistant ne peut pas faire quelque chose, il est peu probable qu'ils réessayent rapidement. Les limites du système sont rapidement perçues par les utilisateurs – avec une petite réserve pour les utilisateurs qui perçoivent uniquement l'assistant comme un moyen de divertissement - en fonction de quoi ils adoptent différents types d'utilisations. Face à la perception de ces limites, certains utilisateurs décident de ne plus utiliser du tout l'assistant :

« Je l'ai utilisé de moins en moins en fait ! Au fur et à mesure que la curiosité s'est estompée j'ai arrêté de l'utiliser. C'est étonnant parce que généralement quand tu as une technologie à ta disposition que ton utilisation s'améliore au fil du temps, parce qu'il y a le temps d'adaptation dans les pratiques. » (Maxime)

Le défi consiste à apprendre aux utilisateurs comment ces systèmes évoluent et s'améliorent au fil du temps. Les utilisateurs ne recherchent généralement pas d'eux-mêmes de nouvelles fonctionnalités. Par contre, ils insistent sur l'importance de la qualité de l'assistant et du smartphone qui le supporte. Nous l'avons vu, le fait d'avoir un bon assistant constitue, selon Budju et Laubheimer (2018), un prérequis essentiel pour répondre à un besoin avec succès. Les utilisateurs n'hésitent d'ailleurs pas à comparer les performances des différents assistants.

« j'ai changé de téléphone récemment et ça marche sûrement mieux du coup je me rends compte que le fait que ça marche mieux intensifie mon utilisation. » (Nathan)

«[A propos de Google Assistant] ça ne marche jamais c'était très chiant mais par contre depuis que je suis repassé à l'iPhone là y a un mois-là, avec Siri ça va nickel quoi. Tous les trucs que je fais avec Siri ça marche (...) J'avais aussi essayé avec Bixby au début sur mon ancien téléphone mais ça j'ai vite abandonné. Même avec les instructions sur internet le truc était incapable de faire ce que je lui disais de faire. C'est largement le plus nul d'après moi, il ne capte rien. » (Sacha)

3.2 Interaction en public

Une préoccupation majeure de nos informants est celle de l'interaction en public. Elle constitue un obstacle à l'utilisation, ou du moins au déroulement d'une interaction sereine. Plusieurs raisons sont évoquées concernant les réticences de l'usage en public. Certains semblent craindre le jugement du fait d'être vu en situation d'interaction vocale avec leur smartphone, soit par principe, soit du fait de l'intrusion que ça représente dans leur vie privée et/ou celle des autres :

« à quel moment est-ce que tu vas accepter le fait que tu parles tout seul à un appareil ? Tu vois moi personnellement j'aurais du mal à accepter cette idée. » (Sébastien)

« [à propos de l'interaction en public] ça me dérange un petit peu... Les autres n'ont pas à connaître toute ma vie et je n'ai pas envie de les déranger non plus. » (Isabelle)

Par ailleurs, interagir en public implique aussi de s'exposer à des bruits parasites qui entravent la compréhension et qui peuvent provoquer des malentendus. Non seulement le bruit modifie la manière d'interagir (répétition, intensité de la voix, articulation, etc.) mais il complique

également la compréhension de la restitution vocale des réponses. De la même manière, un utilisateur mentionne l'impossibilité pour plusieurs usagers d'interagir en même temps :

« Souvent j'ai l'impression qu'avec le bruit, soit il ne va pas comprendre, soit, ce qui arrive encore plus souvent, c'est qu'il comprend autre chose. Et du coup ça peut être super contreproductif parce qu'il va lancer une application ou il va te rediriger vers une page que tu ne voulais pas du tout quoi (...) on n'a pas le contrôle sur le bruit extérieur c'est ça le problème. Si tout le monde demandait à son entourage de se taire pour pouvoir parler à son assistant ce serait l'anarchie ! » (Sébastien)

Vu les préoccupations évoquées ci-dessus, les interactions en public s'avèrent être relativement rares. Or, beaucoup d'adoption aux innovations technologies se font par mimétisme ou par effet de mode. L'adoption de l'assistance vocale ne semble pas faire exception, comme en témoigne un utilisateur :

« Je l'ai vu utilisé une fois par quelqu'un d'autre pour un rappel sur un évènement je me suis dit : c'est super pratique. J'ai testé et je suis devenu adepte de ce truc-là. » (Nathan)

Cependant d'autres utilisateurs regrettent le manque de visibilité que reçoivent les assistants virtuels dans la vie de tous les jours, car ils identifient ce phénomène comme étant un facteur de non-adoption. Selon le processus de domestication de Silverstone (2005), c'est durant la phase de conversion que l'utilisateur retourne temporairement l'objet technologique dans la sphère publique, et que de futurs utilisateurs peuvent comprendre comment il est utilisé. Cette relation avec le monde extérieur aura lieu si un usager raconte ou montre aux autres les différentes fonctionnalités qu'il utilise, mais ce phénomène reste manifestement très occasionnel.

« il y a autre chose c'est que pour que je commence à utiliser quelques choses il faut souvent que je vois une personne l'utiliser. Et je ne connais personne qui utilise un assistant vocal, ou en tout cas je n'en ai jamais vu quoi. » (Maxime)

3.3 À propos du sentiment de contrôle

En utilisant les assistants virtuels pour contrôler notre téléphone, nous abandonnons le contrôle de nos informations, mais cette perte de contrôle s'exprime peut-être autrement. Selon Verbeek (2009), de plus en plus de logiciels nous aident à prendre des décisions ou prennent des décisions pour nous. Ces technologies ont souvent pour effet d'inciter les utilisateurs à prendre une décision ou à adopter un comportement par le biais d'une communication subtile qui masque le rôle actif de la machine dans le processus (Verbeek, 2009). Bien que les assistants sont présentés et perçus dans l'imaginaire collectif comme étant au service de l'utilisateur, ils ont également une certaine emprise sur ce dernier. L'assistant vocal est conçu comme capable de trier et hiérarchiser les informations pour faire des recommandations. Les utilisateurs ne s'y trompent pas, mais leur attitude face à ce constat est assez hétéroclite. Certains, en conscience, se préoccupent de la capacité de décision du système, qui peut orienter le choix :

« s'il trouve la réponse ça correspondra généralement à celle donnée par le premier site de la liste et ce n'est peut-être pas ce que je recherche. Du coup, je lui fais faire plutôt des tâches pour lesquels la réponse est unique ». (Isabelle)

D'autres, au contraire, regrettent la faible capacité du système à faire des recherches et à présenter une réponse fondée sur un jugement. Ils considèrent le fait d'obtenir une réponse unique comme un avantage compétitif sur d'autres modes d'interaction, si tant est que cette réponse correspond bien à leurs attentes :

« J'aimerais bien qu'il puisse avoir genre une capacité de décision. Sinon ce n'est pas très intéressant. On peut aussi aller sur la page du moteur de recherche, mais ce qu'on veut c'est aussi avoir la réponse la plus pertinente possible quoi. » (Sébastien)

Le sentiment de contrôle est une thématique qui touche également à l'utilisation concrète de la machine. Comme décrit précédemment, l'assistant peut être « réveillé » par la voix ou par un bouton d'activation. Là encore, les

opinions sont partagées. Pour certain, l'assistant perd une partie de sa pertinence sans un réveil vocal. En effet, comme nous l'avons vu précédemment, certaines utilisations, notamment en situation de multitâche, impliquent un contrôle total par la voix. D'autres considèrent que l'activation par la voix entraîne non seulement des réveils non désirés, mais qu'elle permet aussi une écoute en continu de l'assistant, dont ils ne connaissent pas bien les répercussions. Configurer le bouton d'activation représente une solution pour certains d'entre eux.

« Je sais juste que ça écoute en continu parce que genre que tu dis "Hey Siri" ou "Hey Google" et bien le téléphone réagit. Je trouve ça un peu flippant d'ailleurs. Parce qu'au début il faut aussi configurer sa voix. Du coup si ça tombe il sait aussi différencier quand c'est toi qui parles ou bien quand c'est quelqu'un d'autre. » (Alix)

« Je préfère déclencher avec un bouton. Dans tous les cas. J'ai fini par désactiver la fonction, qui d'ailleurs était active par défaut, de déclenchement par la voix. Je préfère être tout à fait maître de quand et comment je déclenche l'assistant. » (Maxime)

« (...) il y a un truc qui me dérange c'est quand il s'active tout seul. (...) C'est très désagréable parce que ça peut arriver à n'importe quel moment. Du coup tu peux être en réunion et avoir ce téléphone qui se met à parler tout seul quoi... Et puis non seulement il parle tout seul mais quand il est activé il t'écoute et il peut déclencher une commande ! » (Paul)

Enfin, une utilisatrice exprime le sentiment de perte de contrôle par le prisme de la compréhension du système. Nous gardons le contrôle tant que nous comprenons comment fonctionnent les outils dont nous faisons usage. L'interface invisible représenterait à terme, selon elle, une menace pour la compréhension globale des TIC.

« S'il suffit de dire pour que le téléphone fasse : beaucoup de gens qui avaient l'habitude d'utiliser le téléphone normalement vont arrêter de savoir écrire des SMS, télécharger une application, aller sur internet, etc. et donc ils vont devenir dépendants de cet assistant. Pour moi on est déjà en train de perdre beaucoup de compréhension. » (Isabelle)

3.4 À propos de la conversation utilisateur - assistant

Les points de vue exprimés concernant la conversation avec l'assistant vocal mettent en lumière une situation assez paradoxale. Comme nous l'avons vu, l'attrait pour l'anthropomorphisme pousse les concepteurs à initier des progrès dans le processus conversationnel. Nous avons également mis en avant que l'intérêt porté à la « discussion » dépendait en partie de la conception de l'interaction. Cependant, si on considère les perspectives d'avenirs, plusieurs utilisateurs expriment leur enthousiasme à l'idée d'une interaction plus fluide avec leur assistant, notamment par le biais d'échanges de type « conversationnels ». Ceci dit, dans le même temps, ils émettent souvent des réserves par rapport à l'idée d'une conversation régie par « l'intelligence artificielle ». En outre, ils pensent que la fluidité et « l'humanisation » profiteraient avant tout à un public particulier (notamment les personnes handicapés ou les personnes âgées).

«[...] je ne suis pas certain que j'aurais envie d'avoir une conversation avec un robot mais... d'un autre côté c'est un peu ça que je leur demande aussi, c'est qu'ils deviennent de plus en plus naturels.» (Sacha)

« [...]ce serait cool de pouvoir à la fin, genre juste causer normalement comme si tu parlais à un pote ou quoi. Et je pense que le truc qui bloque pour le moment c'est le fait qu'il ne soit pas fluide.» (David)

«[en parlant de la conversation] Si on prend un exemple extrême... genre dans le cadre du petit vieux qui n'a personne à qui parler et qui peut se mettre à parler avec son assistant vocal alors oui pourquoi pas. » (Juliette)

3.5 À propos de l' (in)compréhension

Les utilisateurs que nous avons interrogés n'ont pas toujours fait la distinction entre la modalité vocale et le concept d'assistance vocale. Pourtant, la conception de l'interaction diffère forcément dans les deux cas. Souvent, nos informants se sont montrés satisfaits de la modalité vocale dans

sa capacité à retranscrire du texte. Dans ce cas précis, que ça soit pour la rédaction de messages ou pour d'autres finalités, l'interaction n'est utile que pour la transcription, pas pour l'activation d'une tâche spécifique. Outre générer des « coquilles » dans le texte, l'incompréhension n'aura aucune conséquence sur la réalisation, ou non, d'une commande. Face à ce constat, les utilisateurs se montrent souvent plus enclin à l'utilisation de la voix comme mode de dictée, plutôt que comme mode d'échange.

« En fait quand c'est pour dicter un message ça c'est hyper facile : tu dis genre : "coucou, virgule, ça va, point d'interrogation, à la ligne, blablabla". Ça fonctionne vraiment bien mais quand tu poses une question, une fois sur deux l'assistant dit qu'il ne comprend pas ou il te répond complètement à côté. » (Alix)

« Moi je n'ai rien contre la reconnaissance vocale par exemple. Je l'utilise beaucoup avec World pour retranscrire des textes ou avec Adope Premiere pour sous-titrer des vidéos. (...) Et là tu n'as pas besoin d'adapter ta manière de parler. Tu n'attends rien en retour si ce n'est que le texte rédigé. » (Maxime)

D'autre part, l'incompréhension avec l'assistant est un thème qui divise nos participants. Tous disent en avoir déjà rencontré, mais la fréquence exprimée est relativement inégale. Assez logiquement, ce sont les utilisateurs les plus fréquents que se plaignent le moins des incompréhensions avec leur assistant. La tendance est moins claire entre les utilisateurs fréquents et moins fréquents en ce qui concerne l'origine perçue de ces incompréhensions. Assez aléatoirement, certains mettent en cause leur manière d'interagir, alors que d'autres pointent plutôt les limites de la technologie elle-même. Dans ce dernier cas, ils regrettent notamment la faible capacité de reconnaissance pour les noms propres ou les mots d'origines étrangères.

« Tu as tendance à marmonner ou à ne pas parler trop fort et à ne pas bien articuler donc finalement l'assistant se trompe dans la traduction » (Isabelle)

« C'est normal que des fois elle [l'assistant] ne puisse pas répondre. Si la réponse n'est pas sur internet ou que t'utilises des mots que genre... elle ne connaît pas et bien... Elle ne pourra pas donner de réponse quoi. » (Alix)

Les incompréhensions nécessitent des adaptations de la part de l'utilisateur qui, nous l'avons vu, peuvent dépendre de la conception du système. La frustration sera ainsi plus ou moins bien gérée par l'utilisateur, en fonction de sa capacité d'adaptation et de résilience. Un utilisateur nous informe qu'il a déjà été jusqu'à invectiver son assistant vocal. Cette pratique interroge d'ailleurs le CNPEN¹² (2020), qui pose la question de savoir s'il s'agit d'un acte moralement répréhensible, en particulier lorsqu'il a lieu en présence d'enfants.

« Parfois, quand il est complètement à côté de la plaque ça m'arrive de l'insulter. Je sais bien que ça ne change rien, c'est plus pour rire qu'autre chose, mais c'est frustrant parfois. » (Sébastien)

« (...) je ne m'énerve pas si ça ne marche pas ! » (Juliette)

Conclusion

1. Implications pour le développement des compétences

La présente étude fournit une nouvelle grille de lecture pour comprendre et évaluer la compréhension de l'interaction avec un assistant vocal. Initialement appliquée à la recherche en éducation, la phénoménographie s'intéresse depuis de nombreuses années, de manière extensive, à une série de disciplines telle que la numération, l'optimisation des moteurs, la programmation informatique, la compétence sur le lieu de travail, etc. Dans ces domaines comme dans d'autres, les résultats ont souvent permis d'intervenir dans le développement d'expériences d'apprentissage, ou, dans le cas d'activité avec un dispositif, ont eu un impact sur l'enseignement du design. Notre étude a ceci de spécifique qu'elle aborde l'usage d'une technologie dont nous pouvons qualifier l'utilisation de « contingente ». L'utilisateur du smartphone n'est nullement tenu de faire usage de l'assistance vocale. Il fera le choix d'en faire usage à une fréquence qui

¹² Comité National Pilote D'éthique du Numérique

correspond à son expérience d'interaction. Notre étude n'a pas pour objectif de convertir qui que ce soit à l'utilisation de cette technologie, et ne vise pas non plus à développer nécessairement chez les utilisateurs une conception plus « complexe ». Cependant, nous avons fait le constat qu'au sein d'une communauté d'utilisateurs, cette technologie n'étaient pas toujours comprise et pouvait, le cas échéant, endiguer la simple perspective d'une utilisation et/ou d'une interaction plus adaptée au besoin de l'utilisateur. Une conception moins complexe de l'interaction peut fermer des portes à une utilisation différente de la technologie, et au développement de compétences nouvelles.

À partir des résultats de notre étude, nous pensons qu'il est possible d'élaborer des formations qui permettent non seulement d'élargir l'horizon de perspectives d'utilisation de cette technologie, mais aussi de nuancer plus globalement les conceptions de l'interaction vocale avec une machine. Ces formations pourraient prendre différentes formes. Ce qui est essentiel, c'est que les utilisateurs puissent, au cours d'un processus, avoir l'occasion de remettre en question et d'étendre leur façon actuelle de vivre leur pratique d'interaction et de développer des compétences. Une façon d'aborder cette question serait d'exposer les étudiants aux modes d'expérience qu'ils pourraient développer. La majorité des concepteurs, suite aux conclusions d'études similaires, ont déjà mis à disposition des utilisateurs des listes de suggestions de commandes, dont certains de nos informants ont mentionné l'usage. Nous proposons par exemple qu'à ces « phrases-types » soient accompagnées des « contextes d'usage-types » pour permettre aux utilisateurs une projection en situation réelle d'interaction.

Par ailleurs, nous pensons qu'il est important d'inciter les utilisateurs à réfléchir à ce qu'ils font. Il ne s'agit pas de donner aux utilisateurs la « bonne » façon d'expérimenter, mais il faut, selon nous, prôner une interaction « active ». Nous avons constaté au cours de nos entretiens à quel point l'utilisateur pouvait être « déconnecté » de sa propre interaction. Le fait de mettre des mots sur des processus inconscients leur a parfois permis d'ouvrir les yeux sur leur propre utilisation. Nous en voulons pour preuve certains

échanges, dans lesquels on perçoit des incohérences ou un décalage entre ce qui est dit en début et en fin d'interview. Nous pensons qu'une interaction active associée à des partages d'expériences entre les utilisateurs peut favoriser le développement de compétences. C'est particulièrement vrai pour cette technologie qui, comme nous l'avons mentionné précédemment, génère peu d'influences mutuelles, de par son usage presque exclusivement privé et individuel. Ainsi, nous estimons que le développement d'un programme visant à développer une « communauté de pratiques » serait bénéfique pour les utilisateurs. Cela pourrait prendre la forme d'ateliers ou de forums d'échange.

Dans une moindre mesure, nous pensons que notre recherche représente une contribution utile pour les chercheurs en UX des assistants vocaux. À partir de nos résultats, ils pourront examiner les effets des conceptions des AV sur l'utilisation, et déterminer quels changements dans la conception du système peuvent avoir un impact sur son utilisation. Les résultats de cette recherche constituent un outil supplémentaire pour tenter de comprendre quelles caractéristiques des assistants vocaux affectent l'expérience de l'utilisateur.

2. Discussion

Les résultats de cette recherche ont permis d'explorer avec succès les expériences des utilisateurs en matière d'interaction avec des AV, qui s'articulent selon 5 conceptions distinctes et au moins 6 thèmes récurrents. Les témoignages recueillis nous ont d'abord permis de confirmer beaucoup d'éléments de compréhension issus de notre revue de littérature (au niveau des types de structures de langage, de l'anthropomorphisme, du nombre restreint de tâches, des préoccupations des utilisateurs, etc.) En outre, les résultats de cette recherche concordent avec l'idée selon laquelle l'interaction avec l'AV est encore souvent considérée, et ce même par les utilisateurs les plus fréquents, comme subsidiaire à des interactions traditionnelles (tactiles et gestuelles). Cela rejoint les constatations de Budju et Laubheimer (2018) selon lesquelles, bien que les gens tirent déjà parti de leurs assistants

intelligents, « l'usabilité » ne permet actuellement pas une adhésion à la hauteur des promesses des concepteurs. Dans leur étude, une analogie avait été faite avec le développement des smartphones ; lorsque les auteurs avaient testé leur « usabilité » en 2000, les résultats étaient très peu probants.

Cependant, nous apportons de la nuance dans ce constat « d'échec ». Nous avons distingué plusieurs catégories de conceptions qui s'apparentent à celle d'un « jouet », d'une modalité alternative ou d'une modalité complémentaire d'interaction. Les limitations dans l'interaction proposées par nos utilisateurs furent multifactorielles : tantôt d'ordre physique ou sécuritaire, tantôt d'ordre psychologique ou sociologique. Dans certains cas, les limitations observées semblent pouvoir être liées, dans une certaine mesure, au « degré » de conception qui correspond à celui de l'utilisateur. Celui qui ne voit en l'assistant qu'une simple alternative n'aura pas le même niveau d'acointance avec son dispositif, et semble rencontrer, de ce fait, plus d'incompréhensions avec lui. De la même manière, une conception « routinière » ou « multitâche » de l'interaction conditionne celle-ci à des circonstances particulières, ce qui fait obstacle à un usage de « proximité » conforme à celui désiré par les concepteurs. Par ailleurs, nous avons observé que la conception pouvait influencer la manière de formuler les commandes ou la fréquence d'utilisation.

Dans d'autres cas, il nous est moins évident de comprendre l'origine des limitations à la lumière de la conception des utilisateurs. Notons d'ailleurs qu'il est parfois délicat de saisir l'état de conception exacte d'un usager. Åkerlind (2005) rappelle à ce propos que la phénoménographie n'a pas pour but d'attribuer précisément une catégorie de description à un utilisateur. Elle fournit un cadre à partir duquel se positionner. Pour ne pas faciliter les choses, nous avons constaté que les utilisateurs eux-mêmes utilisent souvent des qualificatifs pour décrire leur expérience d'interaction qui rentrent en contradiction avec la description de leurs usages concrets.

Une question intéressante est également de savoir dans quelle mesure la variation des conceptions est liée à la formation des utilisateurs et/ou à la durée de leur expérience d'utilisation avec l'AV. Nous ne sommes pas capable, à partir de cette étude, de démontrer des liens forts entre la formation et les 5 conceptions de l'interaction. En revanche, une relation semble bel et bien exister. Un utilisateur (Nathan), nous explique par exemple que son métier intensifie son utilisation, alors qu'une utilisatrice (Alix) porte un regard critique sur la technologie qui se nourrit de ses connaissances en marketing. En ce qui concerne la durée de l'expérience d'utilisation, on peut lier certains aspects de conceptions aux phases de domestications de Silverstone (2005). Conformément à cette théorie, on constate bien, pour les expériences d'interaction récentes, un décalage entre les usages observés (typiquement du divertissement) et les usages préconisés par les concepteurs (typiquement de l'assistance). En outre, bien que davantage de recherches sur le sujet soient nécessaires, nous avons constaté que si l'utilisateur n'allait pas jusqu'à envisager l'interaction avec l'AV comme un processus routinier, c'est aussi parce qu'il n'atteignait pas la « phase d'incorporation », durant laquelle la technologie est supposée s'introduire dans le quotidien de l'utilisateur. Sur cette base, nous émettons quelques réserves par rapport à l'idée soutenue par Shulevitz (2018) et Lemaire et al. (2019), selon laquelle l'assistant vocal, par sa nature d'adaptation à l'utilisateur, s'intègre plus rapidement dans la routine de ce dernier.

3. Limitations et recherches ultérieures

Bien que notre recherche ait été réalisée avec rigueur et dans le respect des pratiques recommandées, il y a place pour l'amélioration.

Dans un premiers temps, il est nécessaire de faire mention des critiques et limitations déjà émises à l'égard de la recherche phénoménographique. Nous présentons succinctement deux d'entre elles, afin que le lecteur puisse réfléchir à l'éthique et à l'applicabilité de cette approche :

La première concerne la partialité de l'approche. La recherche telle que nous l'avons réalisée devrait prendre en compte toutes les identités sociales qui jouent un rôle dans la conceptualisation d'une expérience par les participants (Entwistle, 1997). Nous avons été attentif à certains facteurs, mais ce n'est jamais pleinement satisfaisant, et cela peut avoir des répercussions dans les résultats. On peut en effet toujours estimer qu'il existe une manière plus complète de regrouper les différentes manière de faire l'expérience d'un phénomène. On ne parlera donc pas de bon ou de mauvais résultats mais plutôt de résultats plus ou moins complet. Cependant, Webb (1997) met en garde contre la contrainte de faire apparaître des notions trop « binaires », notamment dans la construction de hiérarchies de compréhension. En ce sens, nous pensons que les notions de faible et forte complexité ne doivent pas être vu comme deux pôles d'un continuum sur lequel se situe nécessairement toutes les conceptions.

Ensuite, les chercheurs phénoménographiques doivent également être conscients de la façon dont leurs propres préjugés implicites se manifestent. Ils risquent constamment d'inciter les participants à manifester le phénomène qu'ils cherchent à identifier (Richardson, 1999). En d'autres termes, les entretiens doivent permettre aux participants de construire et structurer leur propre manière de concevoir les choses, mais la conception de la réalité qui est celle du chercheur peut avoir de l'influence sur les sujets. Une manière de limiter les effets néfastes de cette influence aurait pu être de confronter nos participants aux résultats de la présente recherche, afin qu'ils puissent juger eux-mêmes l'interprétation des résultats.

Concernant nos choix méthodologiques, rappelons que les entretiens ont été réalisés à l'aide d'un script préétabli qui cherchait à sonder nos utilisateurs sur leurs conceptions. Faire émerger des croyances et des idées est un défi de taille pour le chercheur. Une critique qui pourrait être faite à l'encontre de notre approche pourrait se concentrer sur les questions de notre guide d'entretien. Ces dernières peuvent s'avérer plus efficaces pour donner des réponses concernant l'utilisation pratique que pour révéler des conceptions

spécifiques, et ce malgré les efforts d'approfondissement qui furent les nôtres pendant les entretiens. Nous avons constaté certaines faiblesses dans nos questionnaires au fur et à mesure de nos interviews et de notre recherche. Une compréhension plus approfondie et plus rapide de la méthode utilisée et de ses implications aurait permis de cibler notre objet de recherche avec plus de précision. Rétrospectivement, nous pensons que la réalisation d'un ou deux entretiens pilotes aurait été judicieux pour tester notre guide d'entretien. Cette pratique fut négligée, bien que recommandée par les chercheurs phénoménographiques et qualitatifs (Sin, 2010).

Nous pensons important de mentionner également que deux questions issues de notre guide d'entretien se sont avérées relativement peu porteuses. Il s'agit des questions relatives au fonctionnement du système et de l'avenir des AV. Nous avons pensé pouvoir tirer des éléments de conception à partir de la projection de nos sujets, et dans les hypothèses qu'ils pouvaient émettre à l'égard de l'aspect technique ou du futur de cette technologie. Malheureusement, les réponses à ces questions sont souvent sorties du champ de l'étude et nous ont uniquement permis d'ouvrir quelques perspectives.

Ensuite, nous pensons que le choix de l'échantillon de notre enquête peut faire l'objet de discussion. Nous avons tenu à recruter des profils variés sur base d'un questionnaire afin de maximiser les chances de variation dans les conceptions. Bien que le nombre de participants fut assez limité, ce choix méthodologique nous a permis de dresser un aperçu du panorama des perceptions pour des utilisateurs lambda, fréquents comme moins fréquents. Cependant, en ne ciblant pas un public en particulier, nous avons obtenu des données forcément moins spécifiques. Nous estimons que plusieurs études similaires plus ciblées permettraient de nuancer et approfondir les conceptions que nous avons identifiées. Nous pourrions notamment imaginer de circonscrire l'étude à un assistant vocal en particulier, ou bien travailler selon l'âge et/ou le sexe des utilisateurs.

Enfin, une contribution utile pourrait également constituer à explorer si l'analyse qui a été faite concernant les interactions sur smartphone pourrait

également être adaptée à d'autres types d'assistants, par exemple les haut-parleurs domestiques. Il pourrait être intéressant de déterminer en quoi le caractère immobile de ces assistants influence les conceptions des usagers.

4. Conclusion générale

Dans l'ensemble, les principaux résultats de cette exploration qualitative à petite échelle de l'interaction avec les assistants vocaux personnels ont mis en évidence un décalage entre la perception des utilisateurs et la grande mutation promise à l'assistance vocale personnelle. Nous avons observé au moins cinq façons qualitativement différentes dont les utilisateurs comprennent l'interaction avec leur assistant vocal, dont une seule d'entre elles entraîne une modification profonde des habitudes et une utilisation intensive. Cependant, à partir des autres catégories de description, en parallèle avec notre analyse thématique, nous avons pu identifier les principales compétences acquises d'une part, et les freins identifiés par les utilisateurs d'autre part, dans l'expérience interactionnelle. Cette recherche a donné lieu à plusieurs suggestions visant à modifier ou élargir les perceptions. Outre les aspects techniques du système, dont certains réclament encore des améliorations, nous proposons que les concepteurs aident les utilisateurs à contextualiser leurs potentiels usages et nous prônons par ailleurs une interaction « active », motivée par un échange de pratiques entre les utilisateurs. Nous pensons plus globalement que la diversité de perception identifiée, qui n'a pas été largement diffusée auparavant, devrait être prise en compte dans le cadre du développement de nouvelles fonctionnalités, pour permettre aux utilisateurs de tirer un maximum de bénéfices de leur assistant vocal personnel. Il convient toutefois de rappeler que nos résultats sont pertinents à petite échelle et que la recherche qualitative ne doit pas être généralisée à une population d'utilisateurs plus large.

Bibliographie

Abeysekera, I. (2018, juillet 14). *What is phenomenography*. [video]. Youtube. https://www.youtube.com/watch?v=ks_waDFvOGE

Åkerlind, G. (2005). Learning about phenomenography : Interviewing, data analysis and the qualitative research paradigm. In *Doing Developmental Phenomenography* (p. 63-73).

Alain. (2019, avril 21). Quels livres faut-il lire pour comprendre l'affordance ? *Blog Ux*. <https://blog-ux.com/quels-livres-faut-il-lire-pour-comprendre-laffordance/>

Andretta, S. (2012). *Ways of Experiencing Information Literacy : Making the Case for a Relational Approach*. Elsevier.

Ashworth, P., & Lucas, U. (2000). Achieving Empathy and Engagement : A practical approach to the design, conduct and reporting of phenomenographic research. *Studies in Higher Education*, 25(3), 295-308. <https://doi.org/10.1080/713696153>

Barnard, A., McCosker, H., & Gerber, R. (1999). Phenomenography : A Qualitative Research Approach for Exploring Understanding in Health Care. *Qualitative Health Research*, 9(2), 212-226. <https://doi.org/10.1177/104973299129121794>

Bartle, R. (1996). *Hearts, clubs, diamonds, spades : Players who suit MUDs*.

Bellegarda, J. (2013). Large-scale personal assistant technology deployment : The siri experience. *INTERSPEECH*.

Bowden, J. A., & Green, P. (2005). *Doing Developmental Phenomenography*. RMIT University Press.

Brousseau, E., & Rallet, A. (1997). *Le rôle des technologies de l'information et de la communication dans les changements organisationnels*. 18.

Budiu, R. (2019). *Mental Models for Intelligent Assistants*. Nielsen Norman Group. <https://www.nngroup.com/articles/mental-model-ai-assistants/>

Budju, R., & Laubheimer, P. (2018). Intelligent Assistants Have Poor Usability : A User Study of Alexa, Google Assistant, and Siri. *Nielsen Norman Group*. <https://www.nngroup.com/articles/intelligent-assistant-usability/>

Budju, R., & Whinton, K. (2018). What Could an Intelligent Assistant Do for You? A Diary Study of User Needs. *Nielsen Norman Group*. <https://www.nngroup.com/articles/intelligent-assistant-user-needs/>

Burbach, L., Halbach, P., Plettenberg, N., Nakayama, J., Ziefle, M., & Calero Valdez, A. (2019). « Hey, Siri », « Ok, Google », « Alexa ». Acceptance-Relevant Factors of Virtual Voice-Assistants. *2019 IEEE International Professional Communication Conference (ProComm)*, 101-111. <https://doi.org/10.1109/ProComm.2019.00025>

Chun, W. H. K. (2011). *Programmed Visions : Software and Memory*. MIT Press.

Commission National Informatique et Liberté. (2020). « A votre écoute », Exploration des enjeux éthiques, techniques et juridiques des assistants vocaux. https://www.cnil.fr/sites/default/files/atoms/files/cnil_livre-blanc-assistants-vocaux.pdf

Comité national pilote d'éthique du numérique. (2020). *Synthèse des contributions*. https://www.ccne-ethique.fr/sites/default/files/publications/cnpen-contributions-decembre-juillet-v2020_10_01.pdf

Coulet, J.-C. (2010). Mobilisation et construction de l'expérience dans un modèle de la compétence. *Travail et Apprentissages*, 6(2), 181-198.

Cowan, B. R., Pantidi, N., Coyle, D., Morrissey, K., Clarke, P., Al-Shehri, S., Earley, D., & Bandeira, N. (2017). « What can i help you with? » : Infrequent users' experiences of intelligent personal assistants. *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 1-12.
<https://doi.org/10.1145/3098279.3098539>

Davis, F. D. (1993). User acceptance of information technology : System characteristics, user perceptions and behavioral impacts. *International Journal of Man-Machine Studies*, 38(3), 475-487.
<https://doi.org/10.1006/imms.1993.1022>

DeRenzi, B., Dell, N., Wacksman, J., Lee, S., & Lesh, N. (2017). Supporting Community Health Workers in India through Voice- and Web-Based Feedback. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2770-2781.
<https://doi.org/10.1145/3025453.3025514>

Entwistle, N. (1997). Introduction : Phenomenography in Higher Education. *Higher Education Research & Development*, 16(2), 127-134.
<https://doi.org/10.1080/0729436970160202>

Fréjus, M., & Lahoual, D. (2018). Conception d'interactions éthiques et durables entre l'humain et les systèmes d'intelligence artificielle. Le cas de l'expérience vécue des usagers de l'IA vocale. *Revue d'intelligence artificielle*, 32. <https://doi.org/10.3166/RIA.32.417-445>

Gershgorn, D. (2018). *The future of software isn't an app. It's a habit*. Quartz. <https://qz.com/1490440/the-future-of-software-isnt-an-app-its-a-habit/>

Guzman, A. (2017). *Making AI Safe for Humans : A Conversation With Siri* (p. 69-85).

Hayes, P. J., & Reddy, D. R. (1983). Steps toward graceful interaction in spoken and written man-machine communication. *International Journal of Man-Machine Studies*, 19(3), 231-284.
[https://doi.org/10.1016/S0020-7373\(83\)80049-2](https://doi.org/10.1016/S0020-7373(83)80049-2)

- Heitzman, A. (2019). How Popular is Voice Search? *HigherVisibility*. <https://www.highervisibility.com/blog/how-popular-is-voice-search/>
- Jiang et al. (2013, juillet). *How do users respond to voice input errors ? | Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. <https://dl.acm.org/doi/abs/10.1145/2484028.2484092>
- Kanter, R. M. (1983). Frontiers for strategic human resource planning and management. *Human Resource Management*, 22(1-2), 9-21. <https://doi.org/10.1002/hrm.3930220104>
- Katarey, M. (2020, décembre 15). How Does Your Google Assistant Really Work? *The Startup*. <https://medium.com/swlh/how-does-your-google-assistant-really-work-88c64d5dd38d>
- Kathryn Whinton, R. B. (2018). *The Paradox of Intelligent Assistants : Poor Usability, High Adoption*. Nielsen Norman Group. <https://www.nngroup.com/articles/intelligent-assistants-poor-usability-high-adoption/>
- Kennedy, G. D. (1978). *The testing of listening comprehension*. Singapore University Press for SEAMEO Regional Language Centre.
- Kiseleva, J., Williams, K., Jiang, J., Hassan Awadallah, A., Crook, A. C., Zitouni, I., & Anastasakos, T. (2016). Understanding User Satisfaction with Intelligent Assistants. *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, 121-130. <https://doi.org/10.1145/2854946.2854961>
- Krummheuer, G. (2015). *Methods for Reconstructing Processes of Argumentation and Participation in Primary Mathematics Classroom Interaction* (p. 51-74). https://doi.org/10.1007/978-94-017-9181-6_3
- Larsson, S. (1986). *Kvalitativ analys—Exemplet fenomenografi*. Studentlitteratur. <http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-32578>
- LaterClips. (2019, octobre 4). *The New Google Assistant Is Scary Smart*. [vidéo]. Youtube. <https://www.youtube.com/watch?v=2nL4dT2vDc>

Lau, J., Zimmerman, B., & Schaub, F. (2018). Alexa, Are You Listening? Privacy Perceptions, Concerns and Privacy-seeking Behaviors with Smart Speakers. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 102:1-102:31. <https://doi.org/10.1145/3274371>

Lemaire, A., Tubiana, S., Schulte, A., Schwing, C., & Laurent, M. (2019). *La voix monte le son : La révolution naissante des assistants vocaux* (p. 40). Roland Berger.

Limberg, L. (2000). Phenomenography : A relational approach to research on information needs, seeking and use. *Undefined*.
<https://www.semanticscholar.org/paper/Phenomenography%3A-A-relational-approach-to-research-Limberg/5a51c4299f88e9125a67122fe3f44c1950700fa6>

Limberg, L.B. (2008). Phenomenography. In L.M Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 611-614). Thousand Oaks, CA : SAGE Publications.

Lopatovska, I., Rink, K., Knight, I., Raines, K., Cosenza, K., Williams, H., Sorsche, P., Hirsch, D., Li, Q., & Martinez, A. (2019). Talk to me : Exploring user interactions with the Amazon Alexa. *Journal of Librarianship and Information Science*, 51(4), 984-997.
<https://doi.org/10.1177/0961000618759414>

Luger, E., & Sellen, A. (2016). « Like Having a Really Bad PA » : The Gulf between User Expectation and Experience of Conversational Agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286-5297. <https://doi.org/10.1145/2858036.2858288>

Marton, F. (1986). Phenomenography: a research approach to investigating different understandings of reality. *Journal of Thought*, 21(3), 28 - 49.

Marton, F. (1994) *The International Encyclopedia of Education*. Oxford: Permagon.

Marton, F., & Booth, S. (1997). *Learning and Awareness*. Routledge. <https://doi.org/10.4324/9780203053690>

Marton, F., & Pong, W. Y. (2005). On the unit of description in phenomenography. *Higher Education Research & Development*, 24(4), 335-348. <https://doi.org/10.1080/07294360500284706>

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 72-78. <https://doi.org/10.1145/191666.191703>

Nogueira, D. M., Maciel, C., Viterbo, J., & Vecchiato, D. (2017). A Privacy-Driven Data Management Model for Smart Personal Assistants. In T. Tryfonas (Éd.), *Human Aspects of Information Security, Privacy and Trust* (p. 722-738). Springer International Publishing. https://doi.org/10.1007/978-3-319-58460-7_49

Norman, D. (1990). *The Design of Everyday Things : Revised and Expanded Edition*. Basic Books.

Pailleur, F. L., Huang, B., Léger, P.-M., & Sénécal, S. (2020). A New Approach to Measure User Experience with Voice-Controlled Intelligent Assistants : A Pilot Study. In M. Kurosu (Éd.), *Human-Computer Interaction. Multimodal and Natural Interaction—Thematic Area, HCI 2020, Held as Part of the 22nd International Conference, HCII 2020, Copenhagen, Denmark, July 19-24, 2020, Proceedings, Part II* (Vol. 12182, p. 197-208). Springer. https://doi.org/10.1007/978-3-030-49062-1_13

Porter, M. (1990). Competitive Advantage of Nations. *Competitive Intelligence Review*, 1(1), 14-14. <https://doi.org/10.1002/cir.3880010112>

PricewaterhouseCoopers. (2018). *Consumer Intelligence Series : Prepare for the voice revolution*. PwC. <https://www.pwc.com/us/en/services/consulting/library/consumer-intelligence-series/voice-assistants.html>

Qvarfordt, P., Jönsson, A., & Dahlbäck, N. (2003). The role of spoken feedback in experiencing multimodal interfaces as human-like. *Proceedings of the 5th international conference on Multimodal interfaces*, 250-257. <https://doi.org/10.1145/958432.958478>

R. Cowan et al. (2017). « *What can i help you with?* » / *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*.

<https://dl.acm.org/doi/10.1145/3098279.3098539>

Relieu, M., & Velkovska, J. (2020). *Ethnographies des agents conversationnels*.

Richardson, J. T. E. (1999). *The Concepts and Methods of Phenomenographic Research*.

<https://journals.sagepub.com/doi/10.3102/00346543069001053>

Rogers, E. M. (1995). Diffusion of Innovations: modifications of a model for telecommunications. In *Die diffusion von innovationen in der telekommunikation* (pp. 25-38). Springer, Berlin, Heidelberg.

Saint-Dizier de Almeida, V. (2009). La dimension relationnelle des communications. *Activités*, 06(2), Article 2.

<https://doi.org/10.4000/activites.2234>

Saldaña, J. (2012). *The Coding Manual for Qualitative Researchers* (Second edition). SAGE Publications Ltd.

Sandberg, J. (2000). Understanding Human Competence at Work : An Interpretative Approach. *The Academy of Management Journal*, 43(1), 9-25. <https://doi.org/10.2307/1556383>

Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétences* (Montréal: Éditions du Renouveau pédagogique).

Shulevitz, J. (2018). *Alexa, How Will You Change Us?* The Atlantic. <https://www.magzter.com/stories/News/The-Atlantic/Alexa-How-Will-You-Change-Us>

Silverstone, R. (2005). *Media, Technology and Everyday Life in Europe : From Information to Communication*. Routledge.

Sin, S. (2010). Considerations of Quality in Phenomenographic Research. *International Journal of Qualitative Methods*, 9(4), 305-319.

<https://doi.org/10.1177/160940691000900401>

Song, Y. W. (2019). *User acceptance of an artificial intelligence (AI) virtual assistant : An extension of the technology acceptance model* [Thesis]. <https://doi.org/10.26153/tsw/2132>

Suchman, L. A. (2007). *Human-Machine Reconfigurations : Plans and Situated Actions* (\$ {number}nd édition). Cambridge University Press.

Thakur, N., & Han, C. Y. (2018). An approach to analyze the social acceptance of virtual assistants by elderly people. *Proceedings of the 8th International Conference on the Internet of Things*, 1-6. <https://doi.org/10.1145/3277593.3277616>

Tisserand-Barthole, C. (2017). L'avenir de la recherche Web sera t-il thématique ? *Bases & Netsources*. <https://bases-netsources.fr/articles-bases/l-avenir-de-la-recherche-web-sera-t-il-thematique>

Trudeau-Pepin, A. (2020). *Le processus d'adaptation aux assistants personnels intelligents sous une perspective communicationnelle. Le cas appliqué de la Société de la Place des Arts*. [Université catholique de Louvain]. <https://dial.uclouvain.be/memoire/ucl/en/object/thesis%3A27625>

Turunen, M., Melto, A., Hella, J., Heimonen, T., Hakulinen, J., Mäkinen, E., Laivo, T., Soronen, H., & Keskinen, T. (2009). User Expectations and User Experience with Different Modalities in a Mobile Phone Controlled Home Entertainment System. *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 1-4. <https://doi.org/10.1145/1613858.1613898>

Université de Sherbrooke. (2013). IMN638 - Interactions visuelles numériques - Chapitre 3 - Interaction Homme-Machine [Diapositives]. https://chamberm.github.io/638/m_chapitre3_1.pdf

Velkovska, J., Zouinar, M., & Veyrier, C.-A. (2020). Les relations aux machines « conversationnelles ». *Rezeaux*, N° 220-221(2), 47-79.

Verbeek, P. P. C. C. (2009). Ambient Intelligence and Persuasive Technology : The Blurring Boundaries Between Human and Technology. *NanoEthics*, 3(3), 231-242. <https://doi.org/10.1007/s11569-009-0077-8>

Webb, G. (1997). Deconstructing deep and surface : Towards a critique of phenomenography. *Higher Education*, 33(2), 195-212.
<https://doi.org/10.1023/A:1002905027633>

Yang, H., & Lee, H. (2019). Understanding user behavior of virtual personal assistant devices. *Information Systems and e-Business Management*, 17. <https://doi.org/10.1007/s10257-018-0375-1>

Yılmaz, M. B., & Rızvanoğlu, K. (2021). *Understanding users' behavioral intention to use voice assistants on smartphones through the integrated model of user satisfaction and technology acceptance : A survey approach*. <https://www.emerald.com/insight/content/doi/10.1108/JEDT-02-2021-0084/full/html>

Zouinar, M. (2019), Characterizing the hybridity of interaction with domestic voice-enabled speakers in Understanding Hybrid Interactions (panel) IEMCA 2019, Mannheim, Allemagne.

Les progrès technologiques récents, notamment dans le domaine de la reconnaissance vocale et de l'intelligence artificielle, ont permis l'avènement d'une assistance vocale aux caractéristiques nouvelles. Aujourd'hui, l'interaction en langage naturelle avec des systèmes d'information est à la portée du plus grand nombre, par l'intermédiaire des smartphones. Après quelques années de manipulation de ces dispositifs, les utilisateurs se sont forgés des représentations et ont développé des usages particuliers. À l'heure où les concepteurs pensent pouvoir opérer une transition vers une ère de l'assistance, en proposant des réponses toujours plus directes et appropriées à l'utilisateur, pour un nombre grandissant de tâches, nous avons voulu questionner la communication avec ces machines et l'usage qui en est fait. Ce mémoire pose donc la question de comment les utilisateurs conçoivent l'interaction avec leur assistant virtuel dans une culture de plus en plus saturée de technologies et d'interfaces aux modalités d'interaction diverses.

Mots-clefs : Assistant vocal, recherche vocale, utilisateur, interaction homme-machine, phénoménographie