



De l'utilisabilité à l'appropriabilité des assistants vocaux. Étudier les interactions vocales en situation domestique à partir d'une démarche centrée utilisateurs

D Lahoual, M Fréjus

► To cite this version:

D Lahoual, M Fréjus. De l'utilisabilité à l'appropriabilité des assistants vocaux. Étudier les interactions vocales en situation domestique à partir d'une démarche centrée utilisateurs. ERGO'IA 2018, Oct 2018, Bidart, France. hal-01882634

HAL Id: hal-01882634

<https://hal.science/hal-01882634>

Submitted on 27 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

De l'utilisabilité à l'appropriabilité des assistants vocaux. Étudier les interactions vocales en situation domestique à partir d'une démarche centrée utilisateurs

D. Lahoual¹

dounia.lahoual@gmail.com

M. Fréjus¹

myriam.frejus@edf.fr

¹EDF Recherche & Développement – 7 Boulevard Gaspard Monge 91120 Palaiseau

RESUME

En intégrant la sphère privée et domestique des habitants, les assistants vocaux se déploient progressivement pour proposer un pilotage de services autour du confort, de la sécurité, de la santé ou du divertissement. Face à la multiplication récente de ces systèmes d'interaction vocales, l'ensemble des attitudes, des besoins, et des usages provenant des utilisateurs demeurent encore méconnus. De plus, ces systèmes questionnent en profondeur les principes d'utilisabilité et d'acceptabilité compte tenu des nouveaux usages et contextes d'utilisation qu'ils portent. Afin d'apporter des premiers éléments de connaissance, nous étudions à partir d'une revue de la littérature, les retours d'expérience de ces assistants vocaux afin d'identifier les facteurs d'appropriation et les freins à l'acceptabilité de ces dispositifs en situation domestique.

Mots-clés

Interaction vocale, assistants vocaux, situation domestique, expérience utilisateur, appropriabilité.

ABSTRACT

Voice assistants are gradually integrating housings to offer services about comfort, security, health or entertainment. Given the recent multiplication of these voice interaction systems, the set of attitudes, needs, and uses from users are still unknown. In addition, these systems question in depth the principles of usability and acceptability with the emergence of new types of contexts and uses. In order to provide the first elements of knowledge, we present a review of the literature in order to identify the factors making easier or harder the appropriation and the acceptability of these devices in a domestic situation.

Keywords

Vocal interaction, voice assistants, domestic situation, user experience, appropriability.

1. INTRODUCTION

Représentant un outil fondamental dans la communication humaine, la parole demeure une forme d'interaction intuitive, invisible et rapide pour réaliser une diversité de tâches à partir des commandes vocales [26]. Pour ces raisons, le développement des technologies vocales a connu un véritable essor dans divers domaines depuis la fin des années 70 grâce à l'apparition de la microélectronique (domotique, industrie, robotique, etc.). Ces technologies ont permis de combiner les systèmes de reconnaissance vocale à des méthodes de synthèse et de commande vocale [24]. Par ailleurs, ces dernières années ont été marquées par d'autres progrès dans le domaine de l'intelligence artificielle, et cela dans plusieurs secteurs : robotique, véhicules autonomes, maisons intelligentes, traitement de la parole et compréhension du langage naturel, etc. [17, 3, 14]. Et cela, dans l'optique, notamment, d'augmenter les capacités humaines en s'appuyant sur une coopération entre l'homme et la machine [12, 19]. Les progrès de l'IA appliquée à la compréhension de la parole ont favorisé le développement de robots conversationnels (chatbots) et d'assistants vocaux répondant aux requêtes variées des usagers (Siri, Google assistant, Cortana, etc.). D'abord limités à l'utilisation du smartphone, les assistants vocaux dits « intelligents » intègrent dorénavant les habitats avec l'arrivée des enceintes intelligentes telles qu'Alexa d'Amazon ou Google Home [31]. Un assistant vocal intelligent représente un logiciel intégré dans un dispositif tangible, disposant de la faculté de dialoguer avec un humain pour réaliser des tâches sur ordre vocal des utilisateurs (exemples : répondre à des questions en sélectionnant des informations sur le web, rappeler des actions importantes aux utilisateurs, effectuer certaines tâches sur des objets connectés, etc.). Disposant d'une capacité d'apprentissage automatique, ces intelligences artificielles se développent continuellement sans l'intervention des développeurs, apprennent de leurs erreurs, et cela, tout en intégrant des informations sur les utilisateurs afin de proposer des services contextualisés et personnalisés [21, 8].

Si la question de la bonne intégration de la reconnaissance de la parole dans les interfaces et produits grand public n'est pas nouvelle [34], la maturité technologique des assistants actuels questionne l'utilisation réelle qui peut en être faite tant par ses utilisateurs finaux que par les concepteurs, en lien avec les autres formes d'interaction. Nous considérons que pour

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author.
ERGO'IA, 3 - 5 Octobre, 2018, Bidart, France.

développer ces nouveaux systèmes vocaux ou les intégrer à des produits ou services, il importe d'étudier et de comprendre les situations d'utilisation dans lesquelles ils s'insèrent, ainsi que les leviers et les freins à leur appropriation. Proposer des systèmes adaptés à l'activité des utilisateurs en prenant en compte le point de vue de l'humain et de ses activités constitue un des principes fondateurs de la démarche en ergonomie.

Afin de développer des connaissances sur ces nouvelles formes d'interaction vocale, nous proposons dans cet article de réaliser un état des lieux sur l'utilisation des systèmes d'interaction vocale en situation domestique. Cette première étape de veille constitue un préalable dans notre projet de recherche orienté vers la conception de services et de dispositifs vocaux dans l'habitat. Ces éléments récoltés doivent nous permettre de répondre à un ensemble de questions :

- *Quels sont les expériences d'utilisation réussies et empêchées avec ces technologies vocales et assistants vocaux ?*
- *Comment améliorer et optimiser les interactions entre les systèmes d'interaction vocale et les utilisateurs à partir de la connaissance du contexte et de l'activité de ces derniers et en lien avec les autres formes d'interaction présentes ?*

Tout d'abord, nous présentons un état des lieux sur les assistants vocaux en nous focalisant sur les utilisations appréciées, les difficultés rencontrées par les utilisateurs, ainsi que les formes d'appropriation identifiées. Puis, nous proposons des principes d'interaction à intégrer lors de la conception. Enfin, nous concluons et discutons des résultats en précisant nos perspectives de recherche.

2. ÉTAT DES LIEUX SUR L'ETUDE DES ASSISTANTS VOCAUX ET DES TECHNOLOGIES VOCALES

Pour réaliser notre état des lieux sur les systèmes d'interaction vocale en situation domestique et professionnelle, nous nous sommes appuyés sur une diversité de sources telles que les travaux scientifiques, les enquêtes menées par des instituts spécialisés, ainsi qu'une diversité d'articles et de billets récents sur les pages du web. Parmi nos références bibliographiques, notre webographie constitue une source importante compte tenu du développement des assistants vocaux et de leur intégration récente dans les foyers. Afin de garantir une validité des données récoltées dans cet état de l'art sur les interactions vocales, nous avons mobilisé une triangulation des données visant une multiplication et une combinaison des matériaux récoltés afin d'obtenir une connaissance plus complète et différenciée du même phénomène étudié [9, 32].

2.1. Les assistants vocaux : des expériences d'utilisation réussies et appréciées

Parmi les utilisations identifiées, nous avons retrouvé des expériences appréciées par les utilisateurs que nous avons catégorisées en trois dimensions : les qualités des formes d'interaction, les qualités du système et les effets des

interactions vocales sur l'activité des utilisateurs. Cette section nous permettra d'aborder les problématiques d'utilité et d'utilisabilité afin de saisir de quelle manière les dispositifs répondent aux besoins, aux attentes et aux buts des utilisateurs « *avec efficacité, efficience et satisfaction, dans un contexte d'utilisation spécifié* » [18].

Les qualités de l'interaction : rapidité, multi-modalité, adaptabilité et courtoisie

Parmi les qualités retrouvées dans les interactions vocales, nous retrouvons tout d'abord **l'instantanéité, la facilité et le confort d'utilisation** de mobiliser la parole comme un moyen d'interaction. Ces caractéristiques permettent une rapidité de commande et de contrôle dépassant largement les actions consistant à presser un bouton ou cliquer à l'aide d'une souris. Les retours vocaux sont notamment rapidement compréhensibles, ils évitent toute ambiguïté que nous pourrions rencontrer via les alarmes sonores et les voyants lumineux [39]. De plus, la commande vocale semble avantageuse pour réaliser des actions complexes nécessitant plusieurs touches et de multiples interfaces [41]. Certaines études démontrent une **complémentarité des formes d'interaction** assurant une meilleure acceptabilité du vocal lorsqu'il est associé à d'autres modalités : les utilisateurs apprécient par exemple de combiner le vocal à la manipulation télécommandée selon les contextes d'usage ou de combiner le vocal avec le canal visuel pour enrichir les données récoltées et vérifier la concordance des informations [41, 5].

L'adaptation à l'utilisateur et au contexte ambiant représente également un facteur clé dans l'utilisation du vocal. Utilisé intuitivement, les utilisateurs n'ont guère besoin de connaître la multitude des services et des applications intégrées dans l'assistant vocal. Le haut-parleur intelligent s'ajuste généralement à la requête et au vocabulaire de l'utilisateur tout en continuant à cumuler des connaissances fines sur les utilisateurs cibles, de façon à contextualiser chaque proposition de service [21]. Par ailleurs, l'expérience utilisateur s'en trouve améliorée lorsque les interactions vocales continuent d'être efficaces en situation de bruits de fond tels que les sources de musique [42]. Enfin, certains utilisateurs semblent apprécier le fait de privilégier certaines **formes d'adressage et d'interaction** avec ces interfaces **intégrant une forme de courtoisie**. Contrairement à l'assistant vocal Alexa proposant des formes d'interaction sur le ton de l'injonction (« *Alexa, éteins les lumières* »), Google Home invite davantage à converser, tout en proposant une interaction plus respectueuse (« *Hey Google, lance l'actualité* »). Une utilisatrice attire ainsi l'attention sur l'impact émotionnel des technologies sur l'humain et sur les valeurs sociales véhiculées par certains assistants vocaux : « *Lorsque j'utilise Alexa après la Google Home, je me sens autoritaire (...) Ce n'est pas isolé de la façon dont les hommes considèrent les femmes, mais cela s'étend aussi sur la manière dont les femmes se considèrent entre elles. Les moindres détails d'expérience utilisateur ont de l'importance* » [29].

Les qualités du système : efficience et adaptabilité inter-individuelle

La **concision et l'adaptation des feedbacks** aux requêtes facilitent l'acceptabilité des technologies vocales. En effet, la pertinence et la justesse des réponses apportées par ces interfaces contribue grandement à la réussite de l'expérience utilisateur. Cette expérience positive est renforcée si l'utilisateur utilise son propre vocabulaire, il s'agit d'une

commande vocale libre [41]. Il a d'ailleurs été observé que certains assistants vocaux, tels qu'Alexa, disposent d'un traitement sémantique plus performant que Siri [42] : « *Hey Siri, combien de temps il reste ? ==> J'ai trouvé un article sur le Times. Dois-je te le lire ? (...) Alexa, combien de temps reste-t-il ? ==> Environ 6min et 10 sec* » (op. cit.). Par ailleurs, nous pouvons remarquer que les technologies vocales intègrent naturellement une forme d'interaction pouvant **s'adapter à une grande variété d'utilisateurs**, dont les personnes en situation de handicap ou « déficientes », les personnes âgées, ainsi que les enfants [33, 25, 5].

2.2. Les effets favorables sur l'activité : cumul, continuité et personnalisation

La **capacité à cumuler des tâches** est l'une des conséquences et des opportunités permises par l'interaction vocale qui est grandement plébiscitée par les utilisateurs. Lors de situations contraintes où les mains et le regard sont affairés, les utilisateurs sont ainsi en capacité de réaliser des doubles tâches durant les routines matinales par exemple : « *D'habitude le matin, je mets de la musique avant de sortir du lit. Ou bien, lorsque je sors de la douche avec les mains mouillées, plutôt que d'aller vers téléphone avec les mains humides, je peux juste lancer la musique pendant que je me prépare pour aller travailler* » [37]. Ce cumul de tâche révèle notamment des utilisations qui s'ancrent profondément dans les routines en faisant de la commande vocale un élément incontournable du quotidien : « *Avant même que j'ouvre les yeux, je discute avec Alexa. Éteins l'alarme. Allume la radio. Quel temps fait-il aujourd'hui ?* » [37]. Dans le cours d'activité des utilisateurs et des habitants, nous observons que les technologies vocales favorisent une **mobilité et une continuité de l'expérience**. Ainsi, la possibilité de gérer des commandes vocales depuis n'importe quelle pièce du foyer sans les mains, ni du regard, favorise non seulement une grande mobilité dans l'espace domestique, mais elle permet notamment de soutenir la dynamique et le passage d'une activité à l'autre dans le cours d'action en situation domestique ou à l'extérieur : « *Si je veux continuer mon livre audio pendant que je m'occupe du linge, du tricot ou autre chose qui me demande un certain niveau d'attention, je dirai : Alexa, lit mon livre, pour reprendre là où j'en étais dans la voiture* ». L'utilisatrice n'a ainsi guère besoin de préciser de quel ouvrage il s'agit, car le dispositif comprend la demande implicite [15].

La **création d'un écosystème personnalisé** représente une autre expérience réussie. En effet, ce système constitué par les utilisateurs vise à contrôler un ensemble dispositifs et de services personnalisés via cet instrument pivot représenté par l'assistant vocal. Ce qui semble faire émerger une forme de centralisation des tâches via cette « tour de contrôle » encourageant une variété de requêtes : « *Nous ne faisons pas des choses nouvelles. Nous réalisons juste un peu plus des tâches que nous faisons déjà (...) Mais maintenant c'est plus facile de le faire. Tout est réuni au même endroit* » [37]. Cette structuration en écosystème semble également évoluer vers une **distribution spatiale et inter-individuelle des dispositifs** dans les foyers. Une enquête menée par Edison Research a démontré que 42% des foyers disposant d'un assistant vocal possède au moins deux haut-parleurs intelligents dans leurs foyers [20]. Lorsque l'utilisation est satisfaisante, ces foyers multiplient ces dispositifs afin de partager cette expérience et

ces services avec d'autres membres du foyer et favoriser une continuité des activités dans une diversité de pièces via la commande vocale.

2.3. Les technologies vocales : difficultés et problèmes éthiques soulevés

Nous abordons dans cette section l'ensemble des empêchements et des épisodes d'échec vécus lors des parcours d'usage de ces dispositifs. Les contraintes rencontrées constituent des résultats pertinents à étudier car elles éclairent l'ensemble des freins à l'acceptabilité, c'est-à-dire les raisons pour lesquelles ces technologies ne sont pas « *en adéquation avec les pratiques, les ressources, les objectifs, le système de valeurs et les normes éthiques des utilisateurs pour être acceptable* » [7, 28]. Nous détaillerons ainsi des formes d'interactions imposées, des défaillances du système évoquées, ainsi que des questions éthiques soulevées.

Des formes d'interaction imposées

➤ Empêchement : visibilité, mobilité et capacité d'agir

L'absence d'écran visuel pour soutenir le vocal afin de proposer une plus grande diversité et richesse des données¹ est souvent considéré comme une contrainte [22, 5, 23, 42]. Une étude réalisée par Luria et al. [22], portant sur une variété de dispositifs domestiques (un haut-parleur à commande vocale ; un écran tactile mural ; une application mobile ; un robot social) a démontré l'importance du canal visuel auprès des utilisateurs pour disposer d'informations sur la réussite ou non des commandes réalisées et pour obtenir des informations sur la situation en cours et le contexte de la machine : « *c'était si facile de réaliser les commandes, mais finalement vous ne saviez pas ce qui n'allait pas avec le vocal* ». D'autres empêchements opératoires ont été recensés tels que l'impossibilité pour un utilisateur d'annuler des achats via la commande vocale (l'utilisateur doit passer une application sur smartphone), de paramétrer ou de personnaliser des fonctions via l'interaction vocale exclusivement [48]. **L'absence de mobilité** de ces dispositifs vocaux (la plupart des haut-parleurs intelligents doivent rester branchés électriquement, excepté le modèle Tap d'Amazon par exemple) empêche son transport d'une pièce à l'autre [42, 27]. Cette contrainte peut créer des ruptures dans la continuité des activités lorsque l'habitant se déplace dans le logement.

Les publics âgés mettent notamment l'accent sur certains risques inhérents à la commande vocale, celui de développer une **passivité et une oisiveté** dans les tâches quotidiennes qu'ils préfèrent prendre le temps de réaliser manuellement [39, 33]. Les travaux de Portet et al. [33] sur les systèmes vocaux en situation domestique soulignent l'importance de préserver auprès des personnes âgées ce sentiment de contrôle et leur capacité à agir en toute autonomie : « *J'aime bien agir plutôt que parler (...) J'aime bien fermer mes volets, etc. (...) Je préfère faire les choses parce que sinon c'est glisser vers l'inactivité* » (op. cit.). Il est donc impératif de privilégier des tentatives d'aide à la personne qui ne réduisent ni le sens du contrôle, ni leur propre estime [38].

1. À l'heure où nous écrivons ces lignes, de nouveaux assistants vocaux disposant d'un écran visuel ont été déployés sur le marché. Toutefois, notre étude ne s'appuie pas sur ces dispositifs.

➤ Malaise des usagers : sentiment de présence

L'interaction vocale peut provoquer pour certains utilisateurs une forme de déstabilisation : il s'agit d'un malaise provoqué par un **sentiment de présence** liée à cette parole s'apparentant à une présence humaine : « *Je n'aime pas le fait de me parler à moi-même, ça me donne une impression étrange et très bizarre... C'est un peu effrayant* » ; « *Il pourrait devenir une sorte d'ami au bout d'un certain temps. On pourrait se sentir très à l'aise avec, comme si on était plus seul ? Peut-être que c'est un peu effrayant* » [22]. Ces éléments révèlent ainsi le risque de ces technologies biomimétiques à brouiller les frontières entre l'humain et la machine [1].

Défaillances du système : manque d'intelligibilité, d'efficacité et d'adaptabilité

Parmi les défaillances du système vocal, nous pouvons citer une forme **d'opacité** sur les potentialités des services vocaux. Ce manque d'information peut nuire à l'utilisabilité si les utilisateurs ne connaissent pas l'étendue, ni les limites fonctionnelles de leurs appareils. Pour illustrer ce type de défaillance, nous pouvons citer les défauts de paramétrage de l'Alexa qui ne sont pas immédiatement visibles : le dispositif propose des services ou des produits que l'on ne cherche pas à acheter ; l'utilisateur ne peut pas demander à Alexa d'annuler un achat, il faut le faire depuis une application. Et enfin, Alexa autorise n'importe quelle personne proche du dispositif à commander des produits depuis le compte Amazon de l'utilisateur principal (sans verrouillage de sécurité possible), et cela inclus les enfants, les voix de la radio ou de la télévision [2]. Ce qui induit forcément une perte de contrôle nuisible à l'utilisabilité et à l'acceptabilité de tels dispositifs.

L'échec d'interprétation des commandes et du contexte constitue un autre frein à l'acceptabilité de ces technologies. Parmi cette catégorie nous retrouvons deux principales contraintes. On identifie un problème de fiabilité causé par *mauvaise interprétation* des commandes provenant des biais sémantiques évoqués précédemment ou d'une *mauvaise reconnaissance* pour certaines voix comme celles des personnes âgées, des enfants ou des personnes atteintes d'une maladie [33, 39]. Cet échec de reconnaissance vocale ou de détection des phrases d'activation peut aussi intervenir en cas de perturbation selon l'ambiance sonore tels que les fonds musicaux ou les voix multiples [42]. Enfin, nous identifions une dernière limite liée à l'utilisation *collective*. Si l'assistant vocal ne dispose pas de la capacité à distinguer les différentes voix du foyer, son partage auprès de plusieurs membres peut empêcher la personnalisation du système pour l'utilisateur cible et ainsi modifier la compréhension du contexte de l'utilisateur [4].

Questions éthiques soulevées : confidentialité, discrétion et privauté

L'utilisation des assistants vocaux invite à questionner la dimension éthique des données. En effet, cette problématique met en exergue la confidentialité et la sécurisation des données, mises à mal par ces technologies qui doivent cumuler une quantité massive de données personnelles pour fonctionner efficacement [4]. Dans un souci d'efficacité du dispositif et de contextualisation des services, les utilisateurs sont encouragés à **renoncer à la confidentialité** totale de leurs données dont on recommande un stockage permanent sur les serveurs afin d'assurer une mémoire des actions et des requêtes [33, 4, 23]. Certains utilisateurs évoquent une

intrusion dans l'intimité liée à l'enregistrement continu des activités et des paroles, un inconfort à diffuser certaines requêtes et l'absence d'un bouton mute à part entière sur certains dispositifs : « *Je n'aime pas parce que j'ai la sensation d'être toujours écouté, et je dois rester prudente sur ce que je lui dis* » [22]. Cette question de la confidentialité est également soulevée en situation d'utilisation partagée d'un assistant vocal au sein du foyer : l'intervention spontanée du dispositif pour proposer un produit ou un service pourrait advenir dans un moment inadapté ou en présence d'autres membres ou de personnes extérieures. Ces situations d'utilisation nous amènent ainsi à distinguer deux types de confidentialité à intégrer en conception : il s'agit de la *discrétion*, renvoyant au souhait de ne pas être entendu des autres habitants lors des interactions vocales. Le second correspond à la *privauté*, suggérant la protection et la sécurisation des données personnelles stockées.

2.4. Quelques indices d'appropriation à l'échelle collective

Cette dernière section vise à illustrer une forme d'activité non prévisible par les concepteurs : il s'agit des formes d'appropriation. Elle apparaît sur une perspective temporelle plus longue et intègre une trajectoire d'utilisation marquée par les histoires, les représentations, ainsi que les épisodes de succès et d'échec des utilisateurs [35, 30]. Cette perspective sur l'activité des habitants permettra d'éclairer l'évolution des formes d'usage et la façon dont ils intègrent le dispositif dans leurs « *corps propre et leur culture propre* » [36, 40].

Un outil pluriel au service des habitants

Les habitants intègrent leur assistant vocal dans leurs routines de façon à ce qu'il puisse répondre à leurs propres préoccupations et logiques d'activités. Selon les foyers, les contextes d'usages et les divers projets portés par les habitants, l'assistant vocal incarne une variété de rôle « *supports* » dans la sphère domestique.

Il peut représenter un **instrument de médiation et de résolution de conflits** : l'assistant vocal est mobilisé comme un médiateur impartial lorsque les membres sollicitent son intervention pour apporter une réponse suite à un débat vigoureux au sein du couple afin de résoudre une question : le dispositif peut ensuite être pris à parti par le « gagnant » du débat : « *Tu vois, Alexa le sait. Elle est de mon côté, pas du tien* » [16]. L'assistant peut être sollicité comme un **instrument de contrôle domestique** : le dispositif vocal est utilisé par une mère de famille pour réchauffer les pièces du foyer, de manière discrète en soirée, et à l'insu de son mari qui avait programmé en amont un thermostat plus frais. Dans ce contexte, l'appareil devient un instrument de contrôle domestique visant à prendre l'ascendant malgré un désaccord de principes au sein du couple : « *Elle préfère une maison accueillante et chauffée tandis que je préfère un foyer plus frais dans lequel tout le monde porte un tas de pulls parce que le chauffage est coûteux et que je paye les factures* » [16]. Une autre façon de s'approprier ce dispositif consiste à le mobiliser comme un **instrument de sécurisation domestique** : il peut être sollicité par certains utilisateurs comme un outil visant la simulation d'une présence dans le foyer à l'aide des fonctions de programmation de routines : « *On a eu plusieurs*

cambriolages dans le quartier dernièrement, et on part pour un week-end en famille (...) J'aimerais bien qu'il y ait au moins de la musique deux ou trois fois par jour, voire allumer la TV de temps en temps » [11]. D'autres utilisations sont davantage orientées vers la communication et les activités des enfants. Nous retrouvons ainsi une sollicitation de l'assistant vocal comme un **instrument de communication et de partage ludique**. Une mère de famille évoque ainsi sa façon de l'utiliser pour faire passer des messages et communiquer de façon originale et ludique avec ses enfants : *« J'avais déjà expliqué que les cookies étaient pour un événement en soirée. Les enfants jetaient des coups d'œil furtifs vers le plat (...) Je n'étais pas d'humeur à me répéter (...) En me tournant vers ma meilleure amie virtuelle, je lui demandais : Alexa, lance le titre "U Can't Touch This (...) J'ai fait une petite danse en pointant les cookies et en répétant le refrain »* [6]. L'assistant vocal peut également servir **d'instrument pour renforcer l'autonomie des enfants** en leur déléguant une petite part de contrôle sur les activités de divertissement permises par l'assistant vocal. Un père de famille émet le projet de programmer en amont l'appareil pour permettre à ses enfants (4 et 5 ans) de lancer les musiques souhaitées ou de jouer avec le dispositif : *« Je ne suis pas convaincu du tout par les appareils dédiés aux enfants, juste bons à diffuser quelques chansons enfantines avec un son affreux (...) Je voulais un truc intuitif pour un enfant, mais qui leur permette d'écouter de la "vraie" musique »* [10].

Un outil devenu un membre du foyer

En complément des rôles « *supports* » attribués aux assistants vocaux, ces derniers vont jusqu'à incarner une identité forte et des rôles indispensables dans la sphère relationnelle et familiale.

Il peut être mobilisé comme un **instrument d'éducation** pour soutenir l'apprentissage et répondre aux diverses questions des enfants : *« Maman, comment on épelle... »* ; *« Demande à Alexa »,* ou *« C'est pas grave. J'aurai juste à demander à Alexa quand on sera rentré à la maison »* [6]. Cet exemple révèle la façon dont ces dispositifs sont susceptibles de soutenir le rôle d'un parent dans l'accompagnement des enfants. L'émergence de ces nouveaux types d'usage se font-ils en faveur ou au détriment de l'éducation des jeunes enfants ? Cette question reste ouverte tout en soulevant un ensemble de considérations éthiques sur les impacts de ces technologies sur le développement des enfants. L'assistant vocal peut notamment être sollicité comme un **instrument d'injonction et de persuasion** pour amener les enfants à obéir plus facilement et rapidement : *« Maman, pourquoi je dois porter un manteau ? »* ==> *« Est-ce que je dois demander à Alexa la température qu'il fait ? »* ==> *« Non c'est bon. Je vais le mettre »*. En contexte nocturne, la mère de famille utilise Alexa pour annoncer aux enfants qu'il est temps de se brosser les dents : *« Je demande à Alexa de lancer la playlist pour aller dormir. La musique classique devient un signe apaisant indiquant qu'il est temps de mettre son pyjama et de se préparer à aller dormir »*. La mère de famille souligne que le dispositif vise à incarner symboliquement auprès des enfants une figure d'injonction jouant le rôle du « *méchant flic* » : *« La chose la plus étrange qui s'est produite provient du fait qu'Alexa soit devenue comme un membre de la famille. Il s'agit d'un autre adulte pouvant répondre à des questions et m'aider à jouer le rôle du méchant flic avec mon rôle du bon flic »* [6].

3. DES PRINCIPES DE CONCEPTION D'INTERACTION A PRIVILEGIER

En nous appuyant sur cette première exploration de la littérature sur les assistants vocaux, nous proposons dix lignes directrices pour alimenter la conception de systèmes d'interaction vocale intégrant les logiques d'utilisation et les besoins des utilisateurs :

Feedback utilisateur : encourager un temps de réponse relativement rapide (entre la commande et le retour vers l'utilisateur) ; privilégier la présence d'un voyant lumineux ou d'une réponse vocale confirmant la réception de la commande vocale ; intégrer une transparence des sources utilisées pour les demandes d'information et les services de recommandation divers.

Mode d'interaction et multimodalité : concevoir l'interaction vocale en complémentarité avec les autres modalités existantes pour favoriser une richesse de commande et de feedbacks (vocal, visuel, télécommande) ; proposer des interactions vocales privilégiant la parole libre.

Adaptabilité : assurer une adaptation du système à partir de l'intégration du vocabulaire de l'utilisateur ; proposer un système qui apprend de ses erreurs ; inclure la possibilité de former le système par l'utilisateur.

Confidentialité et sécurité : intégrer un bouton « mute » à part entière ; assurer une confidentialité des données récoltées et stockées sur un serveur protégé ; privilégier une utilisation multi-utilisateurs (profils, droits d'accès) ; permettre le contrôle, l'exploitation et la modification de ses données personnelles et rendre visible les données recueillies par le système.

Contrôle du système : inclure la possibilité aux utilisateurs de garder ou reprendre la main sur le système en cas de dysfonctionnement ou selon les préférences. Il peut s'agir du débrayage, de l'interaction spontanée ou encore une prévision des automatisations selon le niveau de maîtrise du système (via le mobile, le PC ou par commande vocale). Et surtout il s'agit d'informer de façon continue, anticipée ou instantanée les utilisateurs sur les changements d'états du système via un système d'alertes et de notifications pertinentes.

Personnalisation : offrir la possibilité de modifier les sources utilisées pour les recherches d'information et les services de recommandations ; programmer et personnaliser les services avec l'interaction vocale ; proposer une personnalisation originale des objets connectés et des phrases d'activation ; faciliter l'utilisation de la programmation visant le déclenchement d'actions uniques ou regroupées selon des situations spécifiques avec les commandes IFTTT (« *If This, Then That* ») ; adapter le système aux contextes utilisateurs en intégrant la compréhension de l'implicite.

Qualité de capture sonore : favoriser une reconnaissance vocale selon la diversité des contextes ambiants via l'intégration de micro embarqués, de capteurs et de dispositifs d'annulation d'écho (pour compenser la position de l'utilisateur, l'acoustique de l'habitat et la présence de bruits de fond).

Ubiquité : favoriser une continuité de l'expérience entre les membres du foyer, leurs diverses préoccupations domestiques, les différents espaces empruntés, les nombreux dispositifs utilisés et les autres modalités d'interaction existantes en proposant un écosystème se déployant sur une variété de supports.

Multi-utilisateurs : privilégier un système de reconnaissance des voix au sein d'un même foyer (pour éviter le « détournement » du système par des voix extérieures : TV, radio, etc.) ; proposer un paramètre de contrôle pour l'utilisateur principal afin de garder la main sur ses propres données et déléguer s'il le souhaite une partie ou la totalité du contrôle du dispositif à d'autres utilisateurs.

Dimension humanisante : encourager la dimension humanisante lors des interactions pour les utilisateurs intéressés. Cela consisterait à choisir entre un « *assistant fonctionnel* » ou un « *assistant de compagnonnage* ». Ce dernier choix proposerait une variété de critères et d'attributs que l'utilisateur pourrait ajouter, ajuster ou enlever selon ses préférences : variation de voix ou de timbre, intégration de marques de politesse et/ou d'humour, sélectionner un type de personnalité pour l'IA, expression et reconnaissance des émotions, etc.

4. DISCUSSION ET PERSPECTIVES

En définitive, les systèmes d'interaction vocale n'ont pas vocation à remplacer les autres systèmes et modalités d'interaction. En effet, les utilisateurs préfèrent garder le contrôle de leurs dispositifs en ayant recours à diverses modalités selon la situation d'utilisation et les besoins exprimés. Bien que basique dans les formes d'utilisations déployées, l'interaction vocale vient surtout faciliter et renforcer des utilisations préexistantes dans les routines des habitants afin de poursuivre quatre principales logiques d'utilisation : la recherche d'assistance, le divertissement, le pilotage d'objets connectés, ainsi que l'anticipation et l'organisation de son activité. Ainsi, la synthèse des utilisations et des retours d'expériences démontre bien l'importance de mettre à jour les activités réelles des utilisateurs, leurs préoccupations, ainsi que leurs logiques intrinsèques.

Par ailleurs, l'identification des diverses formes d'appropriation a permis d'éclairer des détournements dans l'usage réalisés par les familles afin d'adapter le dispositif à leurs activités, leurs systèmes de ressources et de valeurs. Ces formes d'appropriation ont été possibles car les assistants vocaux remplissaient d'une part, des critères d'**adaptabilité** en s'adaptant efficacement aux situations et aux capacités d'agir des habitants comme l'illustre les situations où les habitants s'en servent comme outil de médiation (la réponse à leurs questionnements est instantanée et ne nécessite pas de recherches fastidieuses sur de multiples interfaces) ou comme outil de contrôle énergétique (la commande vocale favorise un usage rapide et aisé pour gérer la température, elle favorise ainsi une interaction moins rébarbative et moins complexe qu'avec un programmateur classique). D'autre part, les assistants vocaux remplissaient des critères d'**évolutivité** [13] en proposant des fonctionnalités renouvelées en réponse à l'évolution des besoins et des capacités des acteurs. En effet, nous avons pu observer la manière dont l'assistant vocal avait été volontairement intégré dans une sphère familiale et

éducative habituellement occupée par la cellule parentale. Par ailleurs, des mères de famille ont intégré ces dispositifs comme des ressources flexibles s'adaptant à divers contextes domestiques éducatifs, tout en renforçant les pratiques éducatives antérieures. Ces résultats illustrent donc bien l'**appropriabilité** des assistants vocaux pour certains foyers, c'est-à-dire, que ces dispositifs s'adaptaient de façon pertinente aux contextes, aux préférences et au développement des utilisateurs sur le long terme [7].

En complément à l'analyse de la littérature ici présentée, nous réalisons actuellement des observations en situation réelle afin de modéliser le développement et les utilisations d'écosystèmes d'objets connectés, dont des systèmes à base de reconnaissance vocale. Ces travaux de recherche en ergonomie contribuent à la conception de services et de dispositifs soutenant la maîtrise de l'énergie dans l'habitat adaptés au grand public et s'insérant de façon fluide et durable dans les activités et les routines des habitants.

BIBLIOGRAPHIE

- [1] Alliance des Sciences et Technologies du Numérique (ALLISTENE). *Éthique de la recherche en robotique*. Rapport n°1 de la CERNA, Commission de réflexion sur l'Éthique de la Recherche en sciences et technologies du Numérique d'Allistene, 2014.
- [2] S. Cobb, Amazon Echo and the Alexa dollhouses: Security tips and takeaways, 2017. *We live security.com*, <https://www.welivesecurity.com/2017/01/10/amazon-echo-alexa-dollhouse-security/>
- [3] Commission Nationale Informatique & Libertés (CNIL), *Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle*. Synthèse du débat public animé par la CNIL dans le cadre de la mission de réflexion confiée par la loi pour une république numérique, Décembre 2017.
- [4] B. Davis, The problem with voice user interfaces like Amazon Alexa, 2016. *Econsultancy*, <https://econsultancy.com/blog/68499-the-problem-with-voice-user-interfaces-like-amazon-alexa>
- [5] B. DeRenzi, N. Dell, J. Wacksman, S. Lee & N. Lesh, Supporting Community Health Workers in India through Voice- and Web-Based Feedback. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 2017.
- [6] A. B. Donahue, Auntie Alexa: How my Amazon Echo became a part of the family. *Charlottetive.com*, 2017, <https://www.charlottetive.com/auntie-alexa-amazon-echo-became-part-family/>
- [7] M. Dubois & M. E. Bobillier-Chaumon, L'acceptabilité des technologies : bilans et nouvelles perspectives. *Le travail humain*, 72(4), 305, 2009.
- [8] A. Edip, Siri, Alexa, Google Home, M... Les assistants numériques, révolution ou gadget dangereux ?, 2017. *Capital.fr*, <https://www.capital.fr/polemik/siri-alexa-google-home-m-les-assistants-numeriques-revolution-ou-gadget-dangereux-1230739>
- [9] M-F. Fortin. *Le processus de la recherche : de la conception à la réalisation*. Montréal : Décarie éditeur, 1996.
- [10] Forum Hardware, *Topic unique : Assistants intelligents (Amazon Echo/Google Home)*, P.18, 2017 (a) https://forum.hardware.fr/hfr/electroniquedomotiquedi/domotique_maison_connectee/assistants-intelligents-google-sujet_656_18.htm
- [11] Forum Hardware, *Topic unique : Assistants intelligents (Amazon Echo/Google Home)*, P.19, 2017 (b) https://forum.hardware.fr/hfr/electroniquedomotiquedi/domotique_maison_connectee/assistants-intelligents-google-sujet_656_19.htm
- [12] M. Fréjus, *Construire la coopération Homme-IA*. Document interne. Saclay : EDF R&D, 2017.
- [13] M. Fréjus, Élargissement et renouvellement des questions traitées par l'ergonomie dans le domaine du développement durable : retour sur 12 ans de travaux sur les activités domestiques et la maîtrise des consommations énergétiques. *Psychologie française*, à paraître.
- [14] H. Guillaud, Vers une éthique pour l'intelligence artificielle ?, 2016. *Internet actu.net*, <http://www.internetactu.net/a-lire-ailleurs/vers-une-ethique-pour-lintelligence-artificielle/>
- [15] A. L. Hamilton, How i use Alexa, 2017, *Love my Echo.com*, <https://lovemyecho.com/2017/01/03/how-i-use-alexa/>
- [16] S. Heritage, Amazon's Alexa is now part of the family – I just hope she doesn't replace me, 2017, *The*

- [guardian.com,https://www.theguardian.com/technology/2017/nov/19/amazon-alexa-voice-recognition-2017-best-gadgets](https://www.theguardian.com/technology/2017/nov/19/amazon-alexa-voice-recognition-2017-best-gadgets)
- [17] Institut National de Recherche en Informatique et en Automatique (INRIA), *Intelligence Artificielle. Les défis actuels et l'action d'Inria*. Livre blanc N°01. Coordonné par B. Braunschweig, 2016.
- [18] ISO 9241-11, *Exigences ergonomiques pour travail de bureau avec terminaux à écrans de visualisation (TEV)* - Partie 11: Lignes directrices relatives à l'utilisabilité, 1998.
- [19] L. Janin, L'intelligence artificielle : mythes et réalités, 2017. *Stratégie.gouv.fr*, <http://www.strategie.gouv.fr/point-de-vue/lintelligence-artificielle-mythes-realites>
- [20] B. Kinsella, 42 Percent of amazon echo owners have 2 or more devices, 2017. *Voicebot.ai*, <https://www.voicebot.ai/2017/06/22/42-percent-amazon-echo-owners-2-devices/>
- [21] F. Lefevre, Interaction vocale & intelligences artificielles : état des lieux et opportunités, 2017. *Medium.com*, <https://medium.com/@francklefevre/interaction-vocale-intelligences-artificielles-%C3%A9tat-des-lieux-et-opportunit%C3%A9s-3b45239ee4a7>
- [22] M. Luria, G. Hoffman & O. Zuckerman, Comparing Social Robot, Screen and Voice Interfaces for Smart-Home Control. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17). ACM, New York, NY, USA, pp. 580-628, 2017.
- [23] O. Metwally & B. Gregory B, Alexa, start living healthy! Workshop on Interactive Systems in Health Care (WISH), 2017, http://wish.gatech.edu/wp-content/uploads/metwally_omar.pdf
- [24] W. Minker & F. Néel, Développement des technologies vocales. *Le travail humain*, vol. 65, no. 3, pp. 261-287, 2002.
- [25] C. Munteanu & G. Penn, Speech-based Interaction: Myths, Challenges, and Opportunities. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, New York, NY, USA, pp. 992-995, 2016.
- [26] C. Nass & S. Brave, *Wired for Speech – How Voice Activates and Advances the Human-Computer Relationship*. MIT Press, 2007.
- [27] N. Nguyen, I Finally Bought An Amazon Echo And This Is What Happened, 2016. *Buzzfeed.com*, https://www.buzzfeed.com/nicolenguyen/amazon-echo-review?utm_term=.ftMWKOVwe#.iDRxV3Wv
- [28] J. Nielsen, *Usability engineering*. Boston: Academic Press, 1993.
- [29] J. Paolino, Google Home vs Alexa. Two Simple User Experience Design Gestures that Delighted a Female User, 2017. *Medium.com*, <https://medium.com/startup-grind/google-home-vs-alexa-56e26f69ac77>
- [30] C. Pelletier & É. M.-F. Moreau, L'appropriation des technologies de l'Internet et ses facteurs critiques de succès : un défi de plus pour les PME ? *Revue internationale P.M.E : économie et gestion de la petite et moyenne entreprise*, vol. 21, 2, pp. 75-117, 2008.
- [31] P. Pestanes & B. Gautier, Essor des assistants vocaux intelligents : nouveau gadget pour votre salon ou fenêtre d'opportunité pour rebattre les cartes de l'économie du web ?, *Wavestone*, 2017.
- [32] R. Pinard, P. Potvin & R. Rousseau. Le choix d'une approche méthodologique mixte de recherche en éducation. *Recherches qualitatives*, Vol. 24, pp.58-82, 2004.
- [33] F. Portet, M. Vacher & S. Rossato, Les technologies de la parole et du TALN pour l'assistance à domicile des personnes âgées : un rapide tour d'horizon (Quick tour of NLP and speech technologies for ambient assisted living) [in French]. *JEP-TALN-RECITAL 2012, Atelier ILADI 2012: Interactions Langagières pour personnes Agées Dans les habitats Intelligents*, Jun 2012, Grenoble, France. pp.3-16, 2012.
- [34] Portolan, N., Nael, M., Renoullin, J.L. & Naudin, S. (1999). Will we speak to our TV remote control in the future? In: *Proceedings of the 17th international symposium on human factors in telecommunication*, HFT'99, Copenhagen.
- [35] S. Proulx, Trajectoires d'usages des technologies de communication : les formes d'appropriation d'une culture numérique comme enjeu d'une société du savoir. *Annales des télécommunications*, 57(3-4), pp. 180-189, 2002.
- [36] P. Rabardel. *Les hommes et les technologies, une approche cognitive des instruments contemporains*. Armand Colin, Paris, 1995.
- [37] Radiocentre, Getting vocal. How voice-activated devices are increasing radio listening & elevating audio branding, 2017. Full report, http://www.radioplayer.co.uk/sites/default/files/getting_vocal_-_final_report.pdf
- [38] J. Rodin, Aging and health: effects of the sense of control. *Science*, 233(4770): pp.1271- 1276, 1986.
- [39] H. Soronen, M. Turunen & J. Hakulinen, Voice Commands in Home Environment - a Consumer Survey. In *Proceedings of Interspeech 2008*, pp. 2078-2081, 2008.
- [40] J. Theureau. *Appropriation 1, 2, 3 ou Appropriation, Incorporation & « Inculturation »*. Conférence journée Ergo-Idf Appropriation & Ergonomie, 16/06/11, Paris, 2011.
- [41] M. Turunen, A. Melto, J. Hella, T. Heimonen, J. Hakulinen, E. Mäkinen, T. Laivo, H. Soronen, User expectations and user experience with different modalities in a mobile phone controlled home entertainment system. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, New York, NY, USA, 2009.
- [42] K. Whinton, Voice Interaction UX: Brave New World...Same Old Story, 2016. *Nielsen Norman Group.com*, <https://www.nngroup.com/articles/voice-interaction-ux/>