



**HAL**  
open science

# Étude des discours d'éducation aux interfaces vocales : le cas des enceintes connectées

Léonie Béraud

► **To cite this version:**

Léonie Béraud. Étude des discours d'éducation aux interfaces vocales : le cas des enceintes connectées. Sciences de l'information et de la communication. 2018. dumas-03105746

**HAL Id: dumas-03105746**

**<https://dumas.ccsd.cnrs.fr/dumas-03105746>**

Submitted on 11 Jan 2021

**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.

Copyright

## Master professionnel

Mention : Information et communication

Spécialité : Communication Médias

Option : Médias et numérique

# Étude des discours d'éducation aux interfaces vocales Le cas des enceintes connectées

Responsable de la mention information et communication  
Professeure Karine Berthelot-Guiet

Tuteur universitaire : Clotilde Chevet

Nom, prénom : BÉRAUD Léonie

Promotion : 2017-2018

Soutenu le : 21/11/2018

Mention du mémoire : Très bien



## Table des matières

<b>INTRODUCTION</b> .....	<b>5</b>
<b>I – COMMENT PARLER A SON ENCEINTE CONNECTEE ? ANALYSE DE L’EDUCATION D’UN PUBLIC A UNE NOUVELLE INTERFACE</b> .....	<b>17</b>
A. FAUT-IL RE-APPRENDRE A PARLER POUR SE SERVIR D’UNE ENCEINTE CONNECTEE ? .....	18
1. <i>L’ambiguïté du niveau de connaissance technique des cibles</i> .....	20
2. <i>L’émergence d’un langage adapté aux assistants vocaux</i> .....	25
B. FAMILIARISER AUX FONCTIONNALITES .....	30
1. <i>Le problème de l’affordance et des signifiants</i> .....	31
2. <i>La stratégie des marques</i> .....	35
C. HABITUER A L’ABSENCE D’INTERFACE GRAPHIQUE .....	38
1. <i>Chez Amazon : la mise en valeur d’une distance entre utilisateur et technologie</i> .....	39
2. <i>Chez Google : la promesse d’une reconnexion</i> .....	41
CONCLUSION DE LA PARTIE I .....	45
<b>II – DU DISCOURS BIENVEILLANT AUX MISES EN GARDE ETATIQUES : DEUX VISIONS CONTRADICTOIRES SUR LA PLACE DES ENCEINTES CONNECTEES DANS LES FOYERS</b> .....	<b>48</b>
A. L’ENCEINTE CONNECTEE VUE PAR LES MARQUES.....	50
1. <i>Le smart speaker : une gouvernante 2.0</i> .....	50
2. <i>La simulation d’une relation de confiance entre utilisateur et fabricant</i> .....	54
B. L’EMERGENCE D’UN DEBAT SUR LA PROTECTION DES DONNEES PERSONNELLES.....	61
1. <i>En Europe : conséquences de la mise en place du RGPD</i> .....	62
2. <i>Enfants, invités, vie privée : analyse des conseils de la CNIL pour une bonne gestion de son enceinte connectée</i> .....	63
CONCLUSION DE LA PARTIE II .....	69
<b>CONCLUSION</b> .....	<b>72</b>
<b>BIBLIOGRAPHIE</b> .....	<b>77</b>
<b>ANNEXES</b> .....	<b>79</b>

## Remerciements

Je remercie tout d'abord le CELSA et tous mes professeurs pour cette année riche en enseignements. Les compétences acquises au cours de ce Master II me semblent aujourd'hui indispensables pour exercer mon métier de communicante et apporter un regard critique et moderne à mon entreprise.

Je remercie particulièrement Clotilde Chevet qui a suivi ce travail de recherche attentivement et m'a transmis sa passion pour les interfaces vocales avec beaucoup de bienveillance.

Je remercie également Adrien Pitrat, mon rapporteur professionnel, pour sa disponibilité, son regard différent sur mon objet d'étude et ses conseils.

Pour finir, je remercie ma sœur Irina et mes parents qui m'ont très gentiment aidée à relire et peaufiner ce dernier travail universitaire.

## RÉSUMÉ

Ce mémoire analyse les discours sur les enceintes connectées comme Google Home, Echo ou Home Pod, commercialisées depuis 2014 aux États-Unis et 2017 en France. Ces objets sont des enceintes connectées par Internet à un assistant vocal. L'utilisateur peut communiquer avec cet assistant à l'oral, lui poser des questions ou lui donner des ordres. Si les assistants vocaux, comme Siri, sont connus du grand public depuis 2011, les utilisateurs n'avaient jamais été en contact avec eux sans interface graphique.

Nous tenterons de comprendre comment les marques cherchent à développer chez le public des aptitudes gestuelles, des connaissances techniques et une forme de culture numérique à travers leurs stratégies de communication et nous parlerons ainsi d'éducation aux interfaces vocales.

Le deuxième objectif de ce mémoire sera de comparer les discours promotionnels aux prises de parole de l'État. Nous verrons en quoi le discours d'accompagnement aux enceintes connectées de la CNIL nous livre la vision de l'État Français sur les stratégies des GAFAs.

*Mots clés : interfaces vocales / assistants vocaux / éducation / ergonomie / NUI / enceinte connectée*

## INTRODUCTION

Décrypter et imiter le langage humain est un fantasme qui traverse les époques. Au Moyen-Age par exemple, les Hommes inventaient déjà des histoires de têtes de bronze parlantes<sup>1</sup>. Ce n'est que plusieurs siècles après, en 1939, que le laboratoire Bell inventa le premier synthétiseur vocal électronique, le VODER (Voice Operating DEMonstratoR) et présenta en 1952 le dispositif Audrey, capable de reconnaître des chiffres dictés à l'oral, en étant précis à 99%. La première machine parlante avec qui converser est apparu en 1964. Il s'agit du premier chatbot<sup>2</sup> ELIZA, créé par Joseph Weizenbaum, professeur au MIT<sup>3</sup>. La mission de l'agent conversationnel était alors de simuler une séance de psychothérapie.

Pour comprendre comment définir ces technologies imitant le langage humain, il faut d'abord définir l'intelligence artificielle. Nous appelons *intelligence artificielle* l'ensemble des théories et techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine<sup>4</sup>. Le *Natural Language Processing* (NLP) est la branche de l'intelligence artificielle qui permet à un programme de *comprendre* le langage humain. Le NLP s'applique par exemple aux systèmes anti-spams qui signalent les contenus indésirables. Autre exemple, le programme Siri<sup>5</sup> qui transforme la parole en texte, qui est ensuite analysé et génère une action.

Dans les années 80, le NLP connaît des changements structurels grâce à l'introduction des algorithmes de *machine learning* dans le traitement du langage et l'augmentation de la puissance informatique. Selon la définition du site Digital Insiders, le *machine learning*, ou *apprentissage machine* en français, « renvoie à un processus de développement, d'analyse et d'implémentation conduisant à la mise en place de procédés systématiques ». Il s'agit d'une forme de programme permettant à un ordinateur un apprentissage automatisé. De cette façon, la machine peut réaliser un certain nombre d'opérations complexes de manière autonome. L'objectif visé est de rendre la machine capable d'apporter des solutions à des problèmes subtils, par le traitement d'une grande quantité d'informations. L'ordinateur traite les données

---

<sup>1</sup> Mary Branscombe, 26/03/2012, "Reconnaissance vocale : la vie avant Siri" sur Tomsguide.fr <https://www.tomsguide.fr/article/siri-reconnaissance-vocale-histoire,5-118.html>, consulté le 10/06/2018

<sup>2</sup> D'après le site Futura Tech, un chatbot est un programme informatique capable de simuler une conversation avec un ou plusieurs humains par échange vocal ou textuel.

<sup>3</sup> Massachusetts Institute of Technology

<sup>4</sup> Selon la définition du dictionnaire Larousse, sur Larousse.fr, consulté le 10/06/2018

<sup>5</sup> L'assistant vocal d'Apple

et met en évidence les liens entre les différentes situations, et peut ainsi prédire leurs implications. L'intégration du *machine learning* dans le NLP a permis d'accélérer l'analyse de texte et donc d'améliorer la technologie.

L'association du NLP au *machine learning* dans les années 80 permet l'apparition des logiciels de reconnaissance vocale. Cette émergence a permis le développement de cette technologie dans différents domaines : les centres d'appels s'automatisent, Windows s'équipe d'une reconnaissance vocale, les téléphones portables obéissent à certaines commandes vocales simples<sup>6</sup>. La voix s'est donc intégrée progressivement dans plusieurs corps professionnels, puis dans notre quotidien (nos portables, nos jouets, etc).

En 1987, Steve Jobs présente dans une vidéo, le *Knowledge Navigator*, un objet fictif du futur : une tablette tactile parlante<sup>7</sup>. Entre cette présentation précoce de l'assistant vocal et la réelle sortie de Siri en 2011, le NLP a évolué grâce à l'apparition du *deep learning* et du *word embedding*. Le *deep learning*, ou *apprentissage profond* en français, est un sous-ensemble du *Machine Learning*. Il s'agit d'un ensemble de méthodes d'apprentissage automatique qui permet à la machine d'apprendre par elle-même, contrairement à la programmation où elle se contente d'exécuter des consignes prédéterminées. Le *deep learning* est utile car il évite au programmeur d'avoir à assumer les tâches de spécification de fonction (définissant les caractéristiques à analyser à partir des données) et l'optimisation (comment peser les données pour fournir une prévision précise) – l'algorithme fait les deux. Le *word embedding* est une méthode d'apprentissage de représentation de mots. Chaque mot devient un vecteur de nombre, ce qui permet de faciliter l'analyse sémantique, par exemple :

$$\text{Vecteur (roi)} - \text{vecteur (homme)} + \text{vecteur (femme)} = \text{vecteur (reine)}$$

L'évolution des techniques d'apprentissage automatique des machines et de la linguistique informatique ont permis d'améliorer considérablement les technologies NLP. Ainsi, en 2003, un logiciel de reconnaissance vocale étant capable de répondre à son interlocuteur est inventé pour l'armée américaine<sup>8</sup>. Il s'agit alors de l'ancêtre de Siri, l'assistant vocal d'Apple qui sera vendu avec l'iPhone 4S en 2011. Siri imite une

---

<sup>6</sup> Mary Branscombe, 26/03/2012, "Reconnaissance vocale : la vie avant Siri" sur Tomsguide.fr <https://www.tomsguide.fr/article/siri-reconnaissance-vocale-histoire,5-118.html>, consulté le 10/06/2018

<sup>7</sup> Vidéo « Apple future vision 1987 » : <https://www.youtube.com/watch?v=JIE8xk6R11w>

<sup>8</sup> Siri naît à l'origine de la rencontre de plusieurs projets : le projet CALO (Cognitive Assistant that Learns and Organizes) de l'armée Américaine, le projet Vanguard du SRI et le projet Active développé par le SRI et l'Ecole Polytechnique de Lausanne.

conversation et ne limite pas l'utilisateur à une demande ponctuelle mais l'encourage au contraire à faire durer le temps d'interaction et à lui demander tout et n'importe quoi. Depuis, les assistants personnels vocaux ont progressivement vu le jour : Alexa d'Amazon (2014), Cortana de Microsoft (2014), Google Assistant de Google (2016), Bixby de Samsung (2017). Amazon est le premier à associer un assistant personnel vocal à une enceinte et ainsi détacher l'interface vocale de l'écran, deux éléments jusqu'alors complémentaires chez les fabricants.

Il est important de distinguer l'assistant vocal qu'est le logiciel, *software*, comme Siri ou Google Assistant, des enceintes connectées comme Home Pod ou Google Home qui sont les supports, *hardware*. Ce mémoire aura pour objet d'étude les enceintes connectées et l'interface vocale détachée d'un écran. Pour comprendre parfaitement cet objet d'étude, il faut d'abord saisir son fonctionnement technique. Les *smart speakers* vendues en 2018 fonctionnent ainsi : l'enceinte écoute les conversations en attendant l'information qui la déclenche : le *wake-word*, comme "Ok Google" ou "Alexa". Cette fonctionnalité est exécutée en local, il s'agit d'une sorte de bouton on/off qui serait oral. Lorsque la commande est entendue (par exemple, « Quel temps fera-t-il demain à Nantes ? ») l'enceinte se met ensuite à transmettre : elle convertit la voix en texte écrit et envoie la requête à son serveur de référence qui va lui donner la réponse qu'elle retransmet ensuite à l'oral.

La commercialisation de ces enceintes connectées, d'abord par Amazon en 2014 puis par Google et Apple, nous interroge sur l'origine et l'avenir de ces interfaces vocales. L'historique précédent, du VODER aux assistants personnels vocaux, permet de replacer les enceintes connectées au sein d'une évolution et de mettre à distance certains discours médiatiques qui parlent d'une *révolution de la voix*. Comme nous l'avons vu avec le VODER en 1939, la compréhension et l'imitation du langage humain est en étude industrielle depuis le début du XXe siècle et la production d'enceintes connectées s'inscrit dans cette évolution – il ne s'agit pas d'une invention révolutionnaire du XXIe siècle. Le produit semble s'être pourtant adapté au public seulement ces dernières années. Les enceintes connectées auraient atteint un record de rapidité d'adoption. Selon Yann Lechelle, COO de Snips, fabricant d'assistants vocaux français, l'adoption de l'enceinte connectée serait "[...] plus rapide que celle de la voiture, du micro-ondes, de la télévision, du smartphone, de la Xbox<sup>9</sup>". Selon Techcrunch, site américain d'actualités sur les start-up, il aurait fallu 13 ans pour que les

---

<sup>9</sup> Voir entretien avec Yann Lechelle, Annexe 1

télévisions atteignent la barre des cinquante millions d'utilisateurs, quatre ans pour l'accès à Internet, versus deux pour les enceintes intelligentes. Le cabinet indépendant d'étude de marchés IT, Canalys, prévoit 100 millions d'enceintes connectées vendues d'ici la fin 2018<sup>10</sup>. On devine les causes de l'adoption rapide du produit : un climat technophile favorable à son lancement, un système déjà connu des utilisateurs puisqu'il s'agit de logiciels comme Google Assistant ou Siri – des fonctionnalités incluses dans les smartphones Android et Apple depuis 2016 et 2011 – et l'apparence simple du produit. L'essor des enceintes connectées n'est donc pas le fruit d'un hasard mais est bien à l'origine d'une évolution des techniques de reconnaissance et imitation du langage et une adaptation progressive du public à l'assistantat numérique et aux interfaces vocales.

Les enceintes connectées sont aussi arrivées ces dernières années sur les marchés économiques pour plusieurs raisons. En observant les ventes des *smart speakers* au premier trimestre 2018, on distingue trois grands marchés : les Etats-Unis, la Chine et la Corée du Sud. Les fabricants qui arrivent en tête du classement sont Google, Amazon, Alibaba avec T-Mall (équivalent d'Amazon chinois) et Xiaomi<sup>11</sup>. Respectivement ces groupes sont à l'origine, un moteur de recherche, des sites de e-commerce et une entreprise spécialisée dans la téléphonie mobile et l'électronique. Quels ont été les facteurs communs qui ont poussé ces grands groupes à s'aventurer dans la production d'enceinte connectée ? Selon Elena De Vogeleer et Denis Lescop, deux anciens professeurs de l'institut Telecom de Paris Sud, « certaines firmes, à l'instar d'Apple, de Google, de Procter and Gamble, ou encore d'IBM ont compris qu'il leur fallait non pas être des acteurs du marché mais plutôt des supports, pouvant en un sens se substituer au marché dans son acception traditionnelle<sup>12</sup>. ». L'exemple de Google et du marché des applications mobiles illustre bien le propos précédent : en créant le système d'exploitation Android, Google n'est pas devenu un acteur du marché des applications mais bien son support et détient ainsi un pouvoir sur l'existence des applications en décidant si une application apparaîtra sur son catalogue, le Play Store. Selon ces mêmes auteurs, l'émergence d'Internet aurait modifié le jeu concurrentiel. L'objectif n'est plus

---

<sup>10</sup> Canalys, « Media alert: Smart speaker installed base to hit 100 million by end of 2018 », 09/07/2018 – consulté le 02/08/2018

([https://www.canalys.com/static/press\\_release/2018/090718%20Media%20alert%20Smart%20speaker%20installed%20base%20to%20hit%20100%20million%20by%20end%20of%202018.pdf](https://www.canalys.com/static/press_release/2018/090718%20Media%20alert%20Smart%20speaker%20installed%20base%20to%20hit%20100%20million%20by%20end%20of%202018.pdf))

<sup>11</sup> Canalys, « Google beats Amazon to first place in smart speaker market », 24/04/2018 – consulté le 02/08/2018

[https://www.canalys.com/static/press\\_release/2018/Press%20release%20230518%20Google%20beats%20Amazon%20to%20first%20place%20in%20smart%20speaker%20market.pdf](https://www.canalys.com/static/press_release/2018/Press%20release%20230518%20Google%20beats%20Amazon%20to%20first%20place%20in%20smart%20speaker%20market.pdf)

<sup>12</sup> De Vogeleer, Elena, et Denis Lescop. « Plateformes, coordination et incitations », *Management & Avenir*, vol. 46, no. 6, 2011, pp. 200-218.

d'avoir une forte position concurrentielle mais d'être au-delà du marché, comme nous l'avons vu pour Google et les applications. Pour comprendre comment fonctionne l'écosystème du marché des nouvelles technologies, les auteurs citent le normalien et docteur en gestion Olivier Torrès-Blay qui décrit cet écosystème comme “*une coalition hétérogène d'entreprises relevant de secteurs différents et formant une communauté stratégique d'intérêts ou de valeurs structurée en réseau autour d'un leader qui arrive à imposer ou à faire partager sa conception commerciale ou son standard technologique*”. L'objectif est donc de devenir le système d'exploitation et d'imposer un monopole technologique sur un marché afin d'avoir le pouvoir sur l'économie qui en découle et par conséquent maîtriser la filière. Selon Yann Lechelle, “les GAFAs<sup>13</sup> sont en guerre totalitaire car malgré leurs différents marchés, ils cherchent tous à devenir le terminal d'accès. L'interface vocale servirait ainsi à remplir l'espace qui reste<sup>14</sup>.” L'espace qui reste fait référence à ces moments où l'écran, la principale interface, n'a pas sa place : par exemple, quand nous avons les mains prises ou que nous sommes en groupe. Ces instants de déconnexion ont été identifiés par les fabricants d'enceintes connectées et ils ont alors choisi la voix comme l'interface pouvant servir à combler ces vides numériques. Les enceintes connectées s'inscrivent ainsi dans la stratégie décrite par Elena De Vogeleer et Denis Lescop : devenir un support pour maîtriser le *futur* marché de la voix.

Les GAFAs auraient ainsi tout intérêt à nous éduquer à cette *nouvelle* interface. Eduquer vient du latin *ex-ducere*, conduire, conduire hors, élever. Selon le dictionnaire, *éduquer* signifierait « développer chez quelqu'un, un groupe, certaines aptitudes, certaines connaissances, une forme de culture<sup>15</sup> ». Eduquer serait ainsi l'élévation par la culture. Les GAFAs chercheraient donc à nous transmettre leur culture, leur savoir-faire technique pour nous élever et nous permettre d'interagir sans difficulté avec une interface vocale, comme Apple le fit pour l'écran tactile à la sortie du premier iPhone. Rappelez-vous, c'était en 2007, Apple sort l'iPhone 2G, premier smartphone du marché. Les films publicitaires mettent simplement en scène le mobile, une main, et une voix off : “Voilà comment on l'allume” - “Là c'est votre musique” - “ça vos emails”<sup>16</sup>.

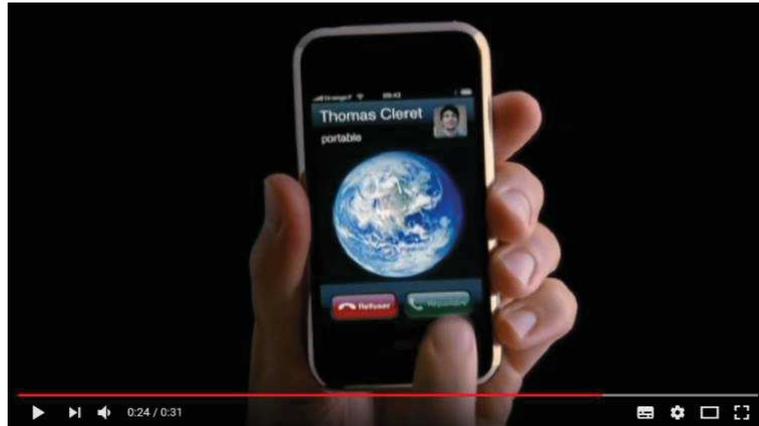
---

<sup>13</sup> GAFAs pour Google, Amazon, Facebook, Apple

<sup>14</sup> Voir entretien avec Yann Lechelle, Annexe 1

<sup>15</sup> Selon la définition du dictionnaire Larousse

<sup>16</sup> <https://www.youtube.com/watch?v=pUfO3dJpcPA>



*Voix off : “Et ça, c’est un appel.”*

*Capture d’écran de la publicité de l’iPhone 2G (sorti en novembre 2007 en France)*

Dès lors, *swiper* avec le pouce, zoomer avec deux doigts, déverrouiller en faisant glisser le doigt de la gauche vers la droite sont devenus des gestes habituels et même naturels. Au-delà de la technique, prendre une photo, lire ses mails ou aller sur Internet sont également devenus des gestes quotidiens que nous avons transférés d’un objet à un autre, de l’appareil photo ou l’ordinateur vers notre téléphone portable. Observe-t-on aujourd’hui un accompagnement éducatif pour l’utilisation des *smart speakers* comme celui que nous avons connu pour les smartphones ? L’utilisation de la voix et la connaissance des assistants vocaux par le public permettent-elles aux marques d’éviter une phase d’apprentissage ?

Nous parlerons ici d’éducation aux interfaces vocales dans la mesure où les publicités et autres discours de marque transmettent un message qui cherche à développer chez le public des aptitudes gestuelles, des connaissances techniques et une forme de culture numérique. Nous verrons comment les discours qui gravitent autour des enceintes connectées cherchent à « développer un sens et une faculté par un entraînement et des exercices appropriés<sup>17</sup> ». Pour comprendre la nécessité d’une éducation aux interfaces vocales, il conviendra de mettre en avant le paradoxe de « l’interface naturelle » qui n’est pas toujours intuitive. Les NUI, Natural User Interface, est un concept né dans les années 90, premièrement conceptualisé par le professeur Steve Mann de l’Université de Toronto. Selon lui, une interface naturelle serait une interface directe, sans métaphore – dont l’utilisation ne demanderait aucune connaissance technique. Un assistant personnel vocal serait ainsi une NUI puisqu’un *smart speaker* fonctionne simplement à la voix. Un utilisateur n’a pas besoin de comprendre le

---

<sup>17</sup> Définition du mot « éducation » du Centre National de Ressources Textuelles et Lexicales

fonctionnement de la machine pour l'utiliser puisqu'il doit simplement faire appel à son propre langage. Paradoxalement, nous verrons en quoi les discours qui accompagnent les enceintes connectées sont avant tout pédagogiques.

Au-delà d'une éducation définie comme un temps de formation et une suite d'exercices, nous étudierons un autre point de vue qui considère l'éducation comme un moyen d'élever par la culture. L'écrivain français André Gide disait : « *Les lois et les morales sont pour l'état d'enfance : l'éducation est une émancipation.* ». Nous analyserons différents discours qui cherchent à émanciper le public en l'aidant à prendre de la distance par rapport à leur relation aux objets numériques.

### CONTEXTE DISCIPLINAIRE

Cette recherche s'articule autour de plusieurs disciplines :

La cybernétique - La cybernétique est « la théorie entière de la commande et de la communication, aussi bien chez l'animal que dans la machine<sup>18</sup> ». Cette discipline interdisciplinaire théorise la relation homme – machine. Nous entendons par relation homme-machine l'interaction d'un utilisateur avec une machine et tous les codes qui régissent cette communication : l'apparence de la machine, ses outils et son langage. Nous serons ainsi amenés à parler d'ergonomie. L'étude ergonomique vise à accroître la productivité d'un travail en adaptant le travail à l'homme. Ce concept se retrouve ainsi lorsque l'on parle d'une technologie ergonomique, il s'agit alors d'une machine qui a été pensée pour son utilisateur. Nous nous servons ici d'ouvrages et d'articles qui s'intéressent aux relations entre utilisateur et objet datant pour la majorité de l'après-guerre, dans les années 50. Si à cette époque, l'informatique n'était pas un outil du grand public, il est justement intéressant de comparer ces écrits qui évoquent nos relations à toute sorte d'objets avec nos habitudes numériques.

L'économie d'Internet – Les écrits sur l'évolution de l'économie mondiale suite au développement de l'utilisation d'Internet nous permettront de contextualiser notre objet d'étude. Comprendre les stratégies et l'environnement économique des fabricants est primordial pour saisir le rôle des enceintes connectées parmi la large palette d'objets

---

<sup>18</sup> Norbert Wiener, *La cybernétique : Information et régulation dans le vivant et la machine*, Seuil, 2014, « Introduction », p. 70

numériques qui nous entourent. Cette mise en contexte nous permettra de mieux comprendre la réaction des États face à la commercialisation des enceintes connectées.

Le droit du numérique – Le droit du numérique interviendra dans la mesure où nous évoquerons les actions des États en réaction au développement des enceintes connectées. En nous basant sur les textes fondateurs de la protection des données personnelles, nous comprendrons ainsi pourquoi les enceintes connectées se sont retrouvées au cœur des préoccupations des instances publiques.

## *PROBLÉMATIQUE*

Ce travail a pour objectif de répondre à la problématique suivante :

*Dans quelle mesure une enceinte connectée, alors qu'elle est censée être une interface naturelle et un support supplémentaire aux assistants vocaux déjà connus du public, nécessite-t-elle plusieurs leviers d'éducation tant sur son fonctionnement que sur sa place dans nos vies et nos sociétés ?*

Pour illustrer nos propos, nous utiliserons à titre d'exemples les modèles d'enceintes connectées existants en 2018 et particulièrement ceux commercialisés en France et sur les marchés américain et anglais (les produits Google Home et Echo d'Amazon). Lorsque nous parlerons d'enceinte connectée, ou *smart speaker*, nous pourrons aussi bien évoquer l'objet en tant que tel que l'assistant vocal qu'il accueille.

## *HYPOTHÈSES*

Ce mémoire a pour objectif de répondre à la problématique précédemment annoncée. Pour cela, notre réflexion s'articulera autour de trois grandes hypothèses :

Notre première hypothèse repose sur le fait que malgré l'essence naturelle de la parole, les enceintes connectées nécessitent une éducation sur leur simple fonctionnement. Cette première hypothèse affirme que les discours des marques sont construits de manière à développer des aptitudes gestuelles et des connaissances techniques chez le public pour accompagner son utilisation des enceintes connectées. Il conviendra de déconstruire les

discours didactiques qui accompagnent la commercialisation des *smart speakers* pour comprendre en quoi ceux-ci nécessitent une *éducation pratique*. Nous mettrons en parallèle de ces messages éducatifs l'apparence intuitive et naturelle des enceintes connectées pour illustrer le paradoxe de ces objets.

Nous verrons dans un deuxième temps comment les marques cherchent à développer une certaine culture du numérique et à transmettre une vision de la place de la technologie dans les foyers à travers leurs stratégies communicationnelles. En communiquant sur la place de l'enceinte connectée dans la maison, sur sa relation avec les enfants ou de sa connectivité avec les autres objets, nous verrons en quoi les discours de marque cherchent à transmettre une culture de l'ultra connectivité.

Notre troisième hypothèse sera d'affirmer que l'État Français considère nécessaire un travail de sensibilisation face à l'arrivée massive des objets numériques dans le quotidien des citoyens. Nous montrerons que l'apparition des enceintes connectées sur le marché français a fortement participé au débat sur la gestion des données personnelles.

Pour une raison de cohérence, nous regrouperons ces deux dernières hypothèses au sein d'une même partie, la troisième hypothèse apparaissant comme une réponse à la seconde. En effet, les discours des marques et de l'État reflètent deux visions qui s'affrontent. L'objectif de cette seconde partie sera de comparer les stratégies de communication des marques et la réponse de l'État pour qui la gestion des données personnelles de ses citoyens remet en cause sa souveraineté.

## *MÉTHODE*

### **Étude d'un corpus de publicités pour Google Home, Alexa et Home Pod**

Nous avons premièrement réalisé une grille d'analyse de films publicitaires de Google, Apple et Amazon datant de 2015 à septembre 2018. Nous avons relevé les caractéristiques suivantes : le marché visé, le thème de la vidéo, les moments représentés, les supports présents (comme la télévision, le smartphone, le livre), les personnages et les voix off, les caractéristiques du dialogue homme – machine (comment les personnages s'adressent aux enceintes, quels tons et quels mots ils utilisent), quelles fonctions de la machine sont utilisées, et les différentes peurs liées aux *smart speakers*.

## **Étude d'un corpus d'articles de presse**

Nous avons également étudié différents articles de presse généraliste et spécialisée. Nous avons relevé ici : la date, l'auteur, les thèmes abordés, les acteurs cités, le point de vue général sur le produit, la description des fonctionnalités et les points de vue des journalistes sur les comportements à adopter avec ce type de technologie (vis à vis de la protection des données personnelles, de l'utilisation des enfants, etc).

## **Étude des prises de parole de la CNIL**

Nous avons analysé différentes pages du site internet de la CNIL afin de comprendre le positionnement de l'instance sur l'émergence des enceintes connectées. L'étude a premièrement débuté avec l'analyse des pages concernant le rôle de la CNIL et les nouvelles normes dues à la mise en place du RGPD puis elle s'est concentrée sur les contenus dédiés aux enceintes connectées et aux préconisations quant à leur utilisation.

Nous analyserons dans une première partie les discours des marques sous l'angle de l'éducation pratique. Pour débiter ce mémoire, nous ferons d'abord le point sur le niveau de connaissances des utilisateurs. Comment les marques considèrent-elles leurs cibles ? Vont-elles immédiatement comprendre dans la mesure où il s'agit d'une *interface utilisateur naturelle* ou sont-elles des débutantes ? Nous nous demanderons en effet s'il ne faut finalement pas ré-apprendre à parler pour utiliser une enceinte connectée. Nous verrons que l'approche pédagogique ne touche pas seulement le langage mais aussi l'absence d'interface graphique, un fait rare pour l'utilisateur. Le design épuré et sans écran des *smart speakers* peut en effet prêter à confusion : que peut faire une enceinte connectée si son aspect ne nous le dit pas ? Ou plutôt, que ne peut-elle pas faire ? Nous verrons qu'au-delà du langage, les marques accompagnent l'utilisateur dans son utilisation pour le guider face à un objet qui ne donne rien à voir.

Notre deuxième partie aura pour objectif de vérifier deux hypothèses : la présence d'une transmission d'une culture numérique par les marques et celle d'un discours de prévention par les états. La première sous-partie sera consacrée à l'analyse des discours de Google, Apple et Amazon. Nous verrons que les GAFAs nous partagent une vision bienveillante de la technologie pour contredire le mythe de la surveillance de masse. Nous verrons en quoi les enceintes connectées sont accompagnées d'un discours plein d'imaginaires et qui qualifie les nouvelles technologies comme notre nouveau compagnon de

vie. En comparaison, nous verrons comment l'État français a agi face à l'apparition des enceintes connectées dans les foyers. Nous introduirons cette seconde partie avec une description du système européen du Règlement Général sur la Protection des Données. En effet, la mise en place de ce texte nous permettra de comprendre dans quelle optique travaille et évolue la CNIL, Commission Nationale de l'Informatique et des Libertés. Nous analyserons ensuite le discours d'accompagnement aux enceintes connectées communiqué sur le site de la CNIL. Que redoute-elle ? Que nous dit sa vision des enceintes connectées sur la nature des relations entre l'État et les GAFA ?

## **I – Comment parler à son enceinte connectée ? Analyse de l'éducation d'un public à une nouvelle interface**

Depuis la sortie d'Echo par Amazon en 2014, les enceintes connectées ont connu un réel succès auprès des consommateurs Américains et deux tiers des acheteurs français se diraient totalement satisfaits de leur acquisition<sup>19</sup>. Pourtant, les enceintes connectées sont finalement peu exploitées. Nous verrons plus tard dans ce mémoire que les utilisateurs se servent principalement de leurs enceintes connectées pour faire ce qu'ils peuvent déjà faire avec leur smartphone ou leur enceinte musicale classique. La seule différence est bien sûr l'utilisation de la voix plutôt que le toucher. Pourquoi les ventes des enceintes connectées ont-elles explosé si leur valeur ajoutée est faible ? Qu'apportent ces objets à leurs utilisateurs ? Si beaucoup de consommateurs utilisent leur *smart speaker* pour mettre de la musique ou demander l'heure, on pourrait penser que c'est surtout le fait de réaliser le rêve de parler à une machine qui les motivent à passer à l'acte d'achat, curieux de voir si cela « marche vraiment ».

Aujourd'hui, pour décrypter les requêtes de leurs utilisateurs, les *smart speakers* fonctionnent ainsi : ils se déclenchent lorsqu'ils entendent un *wake-word*, comme « Ok Google », décryptent la demande vocale de l'utilisateur, la transmettent à un serveur (dans la majorité des cas) et répondent oralement avant de se remettre en veille. Les enceintes connectées actuelles nécessitent ainsi cinq technologies :

- Le *wake-word*, le mot comme “Ok Google” ou “Dis Siri”, qui permet à la machine de déclencher l'enceinte.
- La technologie ASR, *Automatic Speech Recognition*, qui reconnaît un mot à l'oral et le transpose à l'écrit.
- La technologie NLU, *Natural Language Understanding* qui comprend le sens de la phrase. Le NLU est une catégorie des NLP qui tentent de modéliser le langage humain pour comprendre une phrase en fonction de son contexte.
- La partie dialogue : la gestion de la pertinence entre ce qui est demandé et le résultat proposé par le serveur – c'est à dire la possibilité d'apprendre à améliorer les réponses en fonction des interactions précédentes.

---

<sup>19</sup> Selon une étude de l'institut Gfk, 10.2018, citée dans l'article de Salque Alexandre, Brochain Fabrice et François Arias, « Les assistants vocaux ont-ils réponse à tout ? », 01net n°895, numéro du 17 au 30.20.2018

Ce fonctionnement interne est caché derrière un design épuré. La majorité des enceintes connectées n'ont que très peu de bouton voire aucun. Il s'agit généralement d'une enceinte musicale « classique » mais qui serait utilisée avec une télécommande. Hors ici, l'interaction ne doit idéalement passer que par la voix, même pour activer la machine. On lance la machine par le *wake-word*, on lui demande ce que l'on veut, elle s'exécute et se remet en veille, dans l'attente d'un nouvel ordre. Le fonctionnement des enceintes connectées devrait alors être accessible à n'importe qui pouvant parler. À première vue, le temps des notices d'utilisation illisibles et des technologies réservées aux spécialistes nous paraît loin. En effet, quelle interface est plus simple que notre propre corps ? Pourtant, malgré l'essence naturelle de la parole, les discours des marques étudiés montrent qu'un apprentissage du fonctionnement des produits est nécessaire à leur commercialisation. Google et Amazon semblent avoir anticipé un temps d'adaptation pour créer un réel dialogue entre consommateur et enceinte et ont développé un discours pédagogique pour apprendre à l'utilisateur comment parler à l'objet, et ce pour quoi l'utiliser.

Dans quelle mesure, alors que la parole devrait être l'interface la plus intuitive, les fabricants développent-ils des discours pédagogiques sur les enceintes connectées ? Quels procédés communicationnels ont-ils utilisés pour transmettre aux consommateurs les savoir-faire techniques afin qu'ils adoptent rapidement ces objets ?

Cette première partie est consacrée à l'éducation pratique des enceintes connectées et se divisera en deux sous-parties. Nous aurons une première réflexion sur le niveau de connaissances des utilisateurs. Cette première partie nous permettra de comprendre la logique des discours des marques que nous étudierons tout au long de ce mémoire. A partir de nos constats, nous relèverons certaines récurrences présentes dans les publicités qui font ressortir l'apprentissage d'un langage propre à l'enceinte connectée. Notre seconde partie sera consacrée aux fonctionnalités des enceintes connectées. Que font-elles vraiment ? Mais surtout, comment savoir ce qu'elles peuvent faire ?

La troisième partie évoquera l'absence d'interface graphique. Nous analyserons le discours d'accompagnement qui a pour objectif de modifier les habitudes numériques établies au préalable par ces acteurs et qui souhaitent à présent les voir changer.

### **A. Faut-il ré-apprendre à parler pour se servir d'une enceinte connectée ?**

Nous avons vu en introduction que l'interface vocale est arrivée assez tôt dans l'histoire de l'informatique. L'histoire des interfaces ne peut en effet se résumer à l'évolution

de l'écran contrôlé par une souris, vers le tactile puis vers le vocal. L'interface graphique, comme l'interface vocale est née dans les années soixante. C'est en 1963, au MIT, qu'Ivan Sutherland met au point le premier logiciel graphique interactif fonctionnant avec une sorte de stylo. L'interface graphique s'est ensuite progressivement imposée comme l'interface de référence dans l'informatique. En janvier 1983, Apple sort Lisa le premier ordinateur personnel. Bien que la voix intègre rapidement les produits du fabricant américain - en 1984, le Macintosh se présente lui-même, "Hello I am Macintosh and I am glad to be out of the bag."<sup>20</sup> - les nouvelles technologies basées sur une interface graphique se développent plus rapidement et la voix est placée en second plan dans les produits vendus au grand public.

Dans le cadre de son mémoire sur la part de l'imaginaire dans l'interaction avec les assistants personnels vocaux, Clotilde Chevet observe différentes personnes utilisant l'assistant personnel Siri. Nous sommes en 2015 et l'auteure note que la totalité des utilisateurs conservent leur téléphone portable près d'eux, dans leurs mains et ne s'aventurent pas dans un dialogue à distance malgré la possibilité offerte par le smartphone. La conclusion est la suivante : « Nous pouvons supposer que cela s'explique par des raisons pratiques, comme la proximité nécessaire à la lecture, mais nous pouvons aussi supposer que l'usage « télépathique » de la technologie, sans aucun contact physique, n'est pas encore intégré dans les pratiques. ». L'étude montre également que la conversation avec les assistants personnels renforce l'impression d'une conversation en présentiel par l'utilisation de certains mots ou expressions. Les réponses des assistants leur donnent une consistance, une personnalité et presque un corps. Par exemple, si vous demandez à Siri si vous avez un physique ingrat, celui-ci vous répondra : « Bien sûr que non. Et je sais de quoi je parle : je suis aux premières loges à chaque fois que vous utilisez votre iPhone. ». Autre exemple donné dans le mémoire, lorsque vous dites « Bonjour » à votre Galaxy 9, celui-ci pourrait vous répondre « Ah, voilà ma personne préférée. Je suis super contente de te revoir ! »<sup>21</sup>. Déjà, alors que les assistants vocaux sont associés à un écran, les fabricants cherchent à créer une présence pour combler un vide. Aujourd'hui les algorithmes de *machine learning* associent en effet des mots avec des prosodies. Comme dans le langage HTML qui consiste à associer des mots avec des balises (qui servent à la mise en forme d'un texte sur une page web), les algorithmes associent le texte avec des consignes d'émotions. Par exemple, on peut programmer un assistant

---

<sup>20</sup> Lien vers la vidéo : <https://www.youtube.com/watch?v=8bepzUM1x3w>

<sup>21</sup> Clotilde Chevet, "La part de l'imaginaire dans « l'interaction homme-machine » : le cas des assistants personnels numériques", Master 2 Recherche Mention Information et Communication, sous la direction du Pr Karine Berthelot-Guiet, Paris, Université de Paris IV-Sorbonne - CELSA, 2015.

personnel vocal pour qu'il ait une voix enjouée lorsque son utilisateur utilise les mots « blague » ou « rire » et ainsi donner une émotion au dialogue<sup>22</sup>. Pourtant, lors d'un entretien avec Sébastien Fourault, « mobile & Assistant specialist » chez Google, nous évoquons la question de l'imitation d'une voix naturelle et d'une simulation d'une conversation authentique. Google aurait bien comme projet d'imiter au plus près la voix humaine, pourtant Sébastien Fourault nous confiait qu'il n'était pas forcément question de l'utiliser pour le Google Assistant. Alors, quelle relation cherchent à créer les marques entre les utilisateurs et leurs assistants ? Comment créent-ils un lien entre les enceintes connectées et leurs cibles, et dans quel objectif ?

Pour répondre à ces questions, nous nous intéresserons d'abord au niveau de connaissance qu'ont les utilisateurs de leurs enceintes connectées pour comprendre les enjeux des dispositifs de communication des marques. Nous verrons ensuite comment les marques tentent d'éduquer les utilisateurs à un langage propre aux enceintes connectées.

### 1. L'ambiguïté du niveau de connaissance technique des cibles

Alan Cooper<sup>23</sup> a théorisé les degrés de difficulté d'utilisation des interfaces en trois paradigmes. Le premier paradigme, « the technology paradigm » nécessite une connaissance du mécanisme de la machine. L'utilisateur doit comprendre comment le *software* fonctionne pour faire fonctionner la machine. Ce paradigme convient parfaitement aux ingénieurs, curieux de comprendre le mécanisme d'un ordinateur.

*« Engineers prefer to see all of the gears and levers and valves because it allows them to understand what is going on inside the machine<sup>24</sup> »*

Pour le second paradigme, « the metaphor paradigm », l'interface serait compréhensible car expliquée à travers le langage et des pictogrammes. L'auteur introduit ici l'interface graphique moderne destinée aux utilisateurs, avec des fenêtres, des boutons, une souris, des icônes, des métaphores et des menus déroulants. L'auteur conçoit qu'une interface n'est pas un succès grâce à ses métaphores. Par métaphore, nous entendons par exemple la

---

<sup>22</sup> Entretien avec Yann Lechelle, annexe 1

<sup>23</sup> Alan Cooper, «The myth of metaphor», *Visual Basic Programmer's Journal*, 1955

<sup>24</sup> *Ibid.*, page 2 – traduction : « Les ingénieurs préfèrent voir tous les engrenages, leviers et vannes, cela leur permet de comprendre ce qui se passe à l'intérieur de la machine. »

corbeille pour supprimer nos dossiers sur les ordinateurs. Une interface graphique est réussie si celle-ci utilise un vocabulaire restreint pour communiquer avec ses utilisateurs sur un nombre précis d'actions. La métaphore, comme le dossier ou la corbeille, ne fonctionne que si le vocabulaire et les actions demandées sont simples. Les métaphores sont utilisées parce qu'elles sont intuitives. L'auteur explique que nous saisissons le sens d'une métaphore parce que nous la connectons avec un *process* que nous avons déjà pris le temps de comprendre. Nous avons appris plus jeune ce qu'était une poubelle et ainsi nous pouvons facilement rapprocher l'image de la corbeille à la suppression d'un document. Le *paradigme de la métaphore* est une étape supplémentaire dans la relation homme – machine par rapport au *paradigme technique* et s'inscrit comme un lien entre la technique et l'utilisateur.

Le dernier paradigme, « the idiomatic paradigm » repose sur la manière dont nous comprenons les idiomes et les figures de style. Pour Alan Cooper, nous comprenons certaines expressions sans avoir besoin de les assimiler par l'expérience et sans comprendre comment elles fonctionnent vraiment. Une grande partie de ce que nous savons a été appris sans être vraiment compris. Les interactions sociales par exemple, sont acquises sans que l'on sache exactement comment elles fonctionnent. L'auteur fait le rapprochement entre ces apprentissages naturels et l'utilisation de la souris de l'ordinateur. Une souris ne parle pas beaucoup d'elle-même, rien dans son apparence ne précise son fonctionnement et pourtant il est très facile de comprendre rapidement à quoi elle sert. Le paradigme idiomatique fait référence à une interface qui se comprendrait ainsi « naturellement », intuitivement. Les enceintes connectées, parce qu'elles se réfèrent à une pratique naturelle, celle de la parole, se classent-elles dans le paradigme idiomatique ?

L'objectif de cette première partie sera de déconstruire les discours qui vantent l'émergence d'interfaces intuitives pour prouver qu'un discours parallèle d'accompagnement est souvent nécessaire pour familiariser un utilisateur à un nouvel usage. L'iPhone d'Apple en est un exemple. Lorsque la marque a sorti le premier iPhone, l'écran tactile pour téléphone portable existait déjà depuis 1992 – le premier téléphone tactile fut le IBM Simon. Mais en supprimant le stylo, la marque se devait d'expliquer comment l'utilisateur pourrait supprimer, faire coulisser, ou encore zoomer, grâce à un simple mouvement de doigt. Apple a réinventé l'écran tactile et créé une nouvelle interaction homme – machine. Comme nous l'avons vu dans l'introduction, Apple lança une campagne publicitaire ayant pour objectif d'expliquer ce qu'était un iPhone, comment le déverrouiller, utiliser l'écran tactile et manier les différentes applications. En détachant l'interface vocale de l'interface graphique, les fabricants d'enceintes connectées sont-ils en train de révolutionner nos rapports à l'interface vocale que

nous connaissons déjà depuis le développement des assistants vocaux sur smartphone ? Doivent-ils nous accompagner dans cette nouvelle relation ? Mais surtout, que peuvent-ils nous apprendre ?

Dans « Du mode d'existence des objets techniques », Gilbert Simondon distingue deux modes fondamentaux de relation entre l'homme et l'objet technique. L'homme peut être rattaché à l'objet selon un statut de minorité ou un statut de majorité. L'auteur définit le statut de minorité comme celui selon lequel l'objet est un objet d'usage, « [...] nécessaire à la vie quotidienne, faisant partie de l'entourage au milieu duquel l'individu humain grandit et se forme<sup>25</sup> ». Cette relation naît généralement pendant l'enfance et le savoir-faire technique devient ensuite implicite<sup>26</sup>. L'homme sait comment un objet fonctionne sans avoir réfléchi réellement à son fonctionnement, il s'agit ici des habitudes. À l'inverse, « le statut de majorité correspond [...] à une prise de conscience et à une opération réfléchie de l'adulte libre, qui a à sa disposition les moyens de la connaissance rationnelle élaborée par les sciences [...] ». Le statut de majorité fait ainsi référence aux objets plus complexes, apparus plus tard dans nos vies et qui font référence à des codes inconnus ou mal compris, nous poussant à devoir réfléchir à leur fonctionnement.

Dans toutes les vidéos étudiées, les acteurs semblent avoir entre 25 et 40 ans. Il s'agit d'une représentation de la génération née avec Internet, à l'aise avec les nouvelles technologies et surtout familière des interfaces de Google, Apple et Amazon. On parle ici d'une ambiguïté du niveau de connaissance de l'objet dans la mesure où un trentenaire, bien qu'il sache parler voire utiliser un assistant vocal et est globalement familiarisé avec l'univers des fabricants, est obligé d'appréhender l'objet, et de comprendre à quoi celui-ci peut servir avant de l'utiliser. L'enceinte connectée n'appartient certainement pas au paradigme technique dans la mesure où comprendre son fonctionnement n'aide pas forcément l'utilisateur à mieux l'utiliser. Elle n'appartient pas non plus au paradigme de la métaphore parce qu'elle ne comporte aucune interface graphique. Par défaut, nous pouvons considérer qu'elle entre dans la catégorie des paradigmes idiomatique et est donc purement intuitive. Pourtant, son apparence ne donne que très peu d'indices sur le fonctionnement d'un assistant vocal. Pour comprendre ce qui est vraiment intuitif, découpons l'utilisation d'une enceinte connectée en différentes phases :

---

<sup>25</sup> Simondon Gilbert, « Du mode d'existence des objets techniques » Éditions Aubier, Paris, Philosophie, 2012, Chapitre Premier « Les deux modes fondamentaux de relation de l'homme au donnée technique » page 123

<sup>26</sup> *Ibid.*

- Il y a d'abord le *wake-word*. Cette première partie qui est primordiale ne peut pas être intuitive. L'utilisateur doit l'apprendre.
- Le signal lumineux qui confirme à son utilisateur que l'enceinte l'écoute. Ici, ce feed-back est intuitif : on parle et quelque chose se passe – on continue.
- La question : que pouvons-nous demander ? Nous le verrons plus tard, mais cette partie de l'utilisation de l'enceinte connectée est loin d'être intuitive. L'assistant vocal peut servir à un nombre indéfini d'actions et les nombreuses mises à jour ne font qu'augmenter ses capacités.
- La réponse : ici, l'utilisateur est relativement passif.
- La mise en veille : en fonction des enceintes, il y a à nouveau un signal lumineux qui communique à l'utilisateur que c'est la fin de l'échange. Sinon, il le devine.

Sur six phases, deux phases nécessaires à la mise en fonctionnement de l'enceinte ne sont pas intuitives. Pour une personne qui découvre l'interface vocale et qui n'est plus un enfant, sa relation avec une enceinte connectée sera de l'ordre du statut de majorité. Parce qu'il est impossible de deviner le *wake-word* ou les actions précises auxquelles peuvent répondre les assistants, l'utilisateur va devoir réfléchir et anticiper avant chaque interaction avec son enceinte. La vidéo d'une grand-mère appelant son Google Home « Goo-goo »<sup>27</sup> montre que l'enceinte n'est effectivement pas si intuitive – ici l'utilisatrice n'a pas (et ce malgré les recommandations de ses enfants) intégré le *wake-word* et n'arrive donc pas à faire démarrer son enceinte. Cette vidéo qui a connu un grand succès sur la toile, est un bon exemple de ce qu'a montré l'histoire de l'éducation aux objets techniques : l'acquisition de connaissances techniques diffère de l'enfant à l'adulte<sup>28</sup>. Un enfant qui aura vu ses parents dire « Alexa » ou « Ok Google » pourra répéter naturellement et interagir d'une manière plus fluide avec les interfaces vocales. Alors que le produit est assez récent, nous voyons qu'aujourd'hui, ce sont les enfants, qui après avoir compris les étapes de bases, éduquent les adultes au fonctionnement des enceintes connectées<sup>29</sup>.

Malgré des slogans qui prônent la simplicité, comme le « Just ask » d'Amazon, les marques ont bien conscience que leur cible actuelle ne va pas interagir naturellement avec leurs enceintes connectées. Leur utilisation demande un temps d'adaptation et ce ne sera que

---

<sup>27</sup> Lien vers la vidéo : <https://www.youtube.com/watch?v=jEhRaPo2D1M>

<sup>28</sup> Simondon Gilbert, « Du mode d'existence des objets techniques » Éditions Aubier, Paris, Philosophie, 2012, Chapitre Premier « Les deux modes fondamentaux de relation de l'homme au donnée technique »

<sup>29</sup> Entretien avec Yann Lechelle, Snips, voir annexe 1

la prochaine génération, les enfants de la cible actuelle, qui pourront considérer l'interface vocale comme une interface intuitive et développer une relation au statut de minorité avec une enceinte connectée.

Pour Gilbert Simondon, ce sont les artisans et ingénieurs les « témoins et agents de la relation entre la société humaine [...] et le monde des objets techniques [...], ils ont une valeur exemplaire : c'est par eux que l'objet s'incorpore à la culture. ». Les ingénieurs auraient ainsi la responsabilité de créer des objets adaptés à leurs utilisateurs, à leur savoir-faire technique et à leur culture. Les fabricants de *smart speaker* ont ainsi le devoir de transmettre codes et connaissances nécessaires pour faire adopter le produit. Ainsi, comment la communication des fabricants parvient-elle à transmettre ce savoir-faire technique nécessaire au fonctionnement d'une enceinte connectée ? Que tentent d'apprendre les publicités pour enceintes connectées, si ce n'est *parler* ?

Parler nous semble être une pratique *infra-ordinaire*<sup>30</sup>. *L'infra-ordinaire*, expression de George Perec, désigne l'habituel : objets, pratiques et coutumes que nous ne questionnons plus, comme notre sac à main, papier peint, table, rythme de vie... Toutes ces évidences qui composent notre quotidien, tout ce que « nous vivons sans y penser, comme s'il ne véhiculait ni question, ni réponse, comme s'il n'était porteur d'aucune information.<sup>31</sup> ». Dans son texte *L'infra-ordinaire*, l'auteur invite le lecteur à « questionner ses petites cuillères », à « interroger ce qui semble tellement aller de soi que nous en avons oublié l'origine » en d'autres mots, à « fonder [enfin] notre propre anthropologie : celle qui parlera de nous, qui ira chercher en nous ce que nous avons si longtemps pillé chez les autres. Non plus l'exotique, mais l'endotique.<sup>32</sup> ». Les questions de l'auteur, comme « Qu'y a-t-il sous votre papier peint ? », « Combien de gestes faut-il pour composer un numéro de téléphone ? Pourquoi ? », peuvent s'appliquer à la question du langage courant, élément quotidien qui nous semble intuitif. En effet, nous ne questionnons plus nos mots ou notre choix de syntaxe. Pourquoi disons-nous « oui » pour acquiescer et pas « non » ? Pourquoi l'utilisation de la marque « frigidaire » pour parler d'un réfrigérateur nous paraît-elle si normal ? Pourquoi disons-nous « allo » en décrochant le téléphone ? En cherchant à imiter la conversation humaine, les fabricants d'assistants - les dialoguistes en particulier - sont obligés de remettre en perspective

---

<sup>30</sup> Perec Georges, *L'infra-ordinaire*, Editions du Seuil, Paris, Librairie du XXIe siècle, 1989

<sup>31</sup> Perec Georges, *L'infra-ordinaire*, Editions du Seuil, Paris, Librairie du XXIe siècle, 1989, page 11

<sup>32</sup> Perec Georges, *L'infra-ordinaire*, Editions du Seuil, Paris, Librairie du XXIe siècle, 1989, page 12

leurs habitudes langagières et repenser le langage pour déterminer le ton et les mots qui rythmeront les interactions homme-assistant.

Nous verrons dans les prochaines parties dans quelle mesure les marques transmettent certains codes langagiers pour optimiser l'utilisation des enceintes connectées.

## 2. L'émergence d'un langage adapté aux assistants vocaux

### a. La répétition du wake-word

Comme nous l'avons vu précédemment, le *wake-word* et les questions sont les *pain points* des enceintes connectées qui semblent nécessiter un accompagnement.

Le *wake-word*, le « bouton on-off oral » est l'étape du fonctionnement d'une enceinte connectée la moins intuitive. Il est en effet impossible, par la forme de l'appareil, de deviner le mot qui l'allumera. Pour pallier ce manque de clarté, les marques répètent sans cesse le mot magique dans toutes leurs publicités. En effet, dans toutes les vidéos étudiées, les interactions entre une enceinte connectée et son utilisateur sont toujours montrées à partir du *wake-word*. Aucune publicité ne cherche à simuler une interaction authentique en coupant cette partie du *wake-word*. Les marques ne cherchent donc pas à séduire sur ce plan mais bien à éduquer en répétant plusieurs fois le mot déclencheur dans une vidéo afin de s'assurer de la bonne mémorisation du public. Les mises en scène des vidéos valorisent ce premier échange entre l'enceinte et l'utilisateur. En effet, ce mot déclencheur qui peut être un frein à l'utilisation parce qu'il n'est pas naturel, est présenté comme un élément phatique du langage de l'enceinte connectée. La fonction phatique du langage est ce qui précède notre communication ou sert à la prolonger sans communiquer un message<sup>33</sup>. Il s'agit donc ici du « Hey Siri », « Ok Google », « Alexa » ... Les marques montrent que le *wake-word* est essentiel à la conversation avec un assistant. En fonction des situations représentées dans les films des marques, les utilisateurs n'allument pas leur enceinte connectée avec le même ton. Dans une situation banale, les acteurs utilisent un ton enjoué pour énoncer le *wake-word* alors que dans les situations plus dramatiques, le ton va être plus triste ou anxieux. Ainsi, avant même d'avoir commencé à utiliser l'enceinte, rien qu'en l'allumant, les marques créent une

---

<sup>33</sup> Clotilde Chevet, « La part de l'imaginaire dans « l'interaction homme-machine » : le cas des assistants personnels numériques », Master 2 Recherche Mention Information et Communication, sous la direction du Pr Karine Berthelot-Guiet, Paris, Université de Paris IV-Sorbonne - CELSA, 2015 – page 64

relation de complicité entre l'enceinte et l'utilisateur, valorisant ainsi le *wake-word* et l'intégrant de manière plus naturel dans l'échange.

Pourtant, lors de notre entretien chez Google, Sébastien Fourault, « mobile & Assistant specialist » nous confiait que le *wake-word* allait être amené à disparaître. Le fabricant le considère comme une barrière à l'utilisation et diminue l'aspect authentique de la conversation homme – assistant vocal. Selon Yann Lechelle, les fabricants d'enceinte connectée cherchent en effet à tuer ce *wake-word* pour « enlever la friction » et « faire disparaître la technologie ». Cependant, il resterait un effort de contexte à faire. Si une enceinte n'a plus besoin de *wake-word* et écoute en permanence la conversation en attendant de comprendre qu'un utilisateur a besoin d'elle, il faudrait alors qu'elle soit entraînée pour à comprendre l'ambiguïté. L'utilisateur pourrait demander à une autre personne « Il est quelle heure ? » et alors la machine, entendant une question, répondrait automatiquement « Il est onze heures ». Pour le COO de Snips, les enceintes connectées sont face à un problème de contexte et le *wake-word* ne devrait donc pas disparaître bientôt.

b. Donner des ordres et poser des questions : les règles à respecter

Deux actions sont possibles avec les enceintes connectées : on peut leur donner des ordres ou leur poser des questions. Malgré un discours qui prône la simplicité de l'appareil, les fabricants véhiculent à travers leurs publicités, la technique d'un langage propre aux enceintes connectées. Nous verrons plus tard comment la simplicité est devenue le premier argument de vente des marques et comment le fonctionnement des appareils est décrit de manière à faire apparaître une évidence d'utilisation. Cependant, nous verrons ici les règles à respecter pour réussir une conversation avec son enceinte connectée : l'utilisation de l'impératif et d'un pronom interrogatif.

*« This is Google Home. When you get one, you might start to feel like a celebrity with a personal assistant, because you can say stuff like, 'OK, Google, put milk on my shopping list', or, 'turn off the lights'. 'Who do you think you are?' Well, you're just a very important regular person.<sup>34</sup> »*

Voici les premières phrases de la publicité « Google Home : Celebrity », sortie en juin 2017 pour le marché anglophone. Que ce soit chez Google, Amazon ou Apple, dire

---

<sup>34</sup> Extrait de la bande son de la publicité « Google Home : Celebrity », 07.06.2017 – traduction : « Voici un Google Home. Lorsque vous en obtenez un, vous commencez à vous sentir comme une célébrité avec un assistant personnel, car vous pouvez dire des choses comme: «OK, Google, mets du lait sur ma liste de courses» ou «éteins la lumière». "Pour qui vous prenez-vous ?" Eh bien, pour une personne ordinaire très importante. »

« Peux-tu, s'il te plaît, mettre ma playlist 'Summer' » ne fonctionnera pas. Le ton universel du monde des assistants vocaux est l'impératif. La publicité humoristique de Google définit la relation homme – machine : l'utilisateur doit donner des ordres clairs et l'enceinte connectée est là pour obéir. Comme le dit la publicité, l'utilisateur est en position de force vis-à-vis de son assistant, grâce à qui l'on se sent être privilégié. Ce sentiment est créé par l'usage de l'impératif. Cet usage de l'impératif est mis en scène et exagéré dans l'épisode « White people renovating houses » de South Park<sup>35</sup> dans lequel le personnage Cartman prend un plaisir pervers à donner tous types d'ordres à Alexa en prenant un ton très autoritaire. On retrouve ici les imaginaires de la supériorité de la technique face à l'homme et inversement. La peur du remplacement de l'homme par la machine existe depuis la révolution industrielle et l'émergence actuelle de « l'intelligence artificielle » et la peur que celle-ci remplace l'homme s'inscrivent dans ce débat intemporel. Il n'est donc pas étonnant que ces outils ne répondent qu'à l'écoute d'un ordre impératif. Ce ton permet ainsi de replacer la machine comme inférieure à son utilisateur et calme ainsi les peurs communes.

On remarque également que les marques conseillent d'utiliser un pronom interrogatif pour poser des questions en français à son assistant vocal :

Il est recommandé de demander « Quel temps fera-t-il demain ? », « Quelle taille fait la Tour Eiffel ? », « Que veut dire Wacapou ? » plutôt que « Il fera beau demain ? », « Ça veut dire quoi, Wacapou ? ».

Si l'usage de l'impératif est une stratégie pour rassurer l'utilisateur, la nécessité d'utiliser un pronom interrogatif n'est qu'une question de performance de la machine. Les publicités valorisent ce mode de langage pour assurer le meilleur fonctionnement à son utilisateur. On observe d'ailleurs que les fabricants mettent fréquemment leur site à jour pour annoncer les nouvelles capacités des assistants en matière de compréhension de notre langage. Le site d'Amazon liste par exemple les commandes auxquelles Alexa peut répondre pour la seule fonction musicale :

« Play [titre/album/artiste]. »  
« Lis de la musique [émotion]. »  
« Lis [nom de la station]. »  
« Lis [nom playlist]. »  
« Lis une playlist Prime. »  
« Montre-moi des [titres/playlists/genres] disponibles sur Prime. »  
« Qui est le chanteur principal de ce [groupe] ? »

---

<sup>35</sup> Épisode 1 de la saison 21

Ainsi, comme pour la gestuelle apprise pour l'écran tactile, le consommateur enregistre progressivement les différentes techniques pour maîtriser une interface vocale. Si la gestuelle était globalement commune à l'ensemble des applications des smartphones, les formules vocales sont, elles, propres à chaque action et nécessitent ainsi un apprentissage plus lent.

### c. L'apparition d'un lexique propre à la consommation

En analysant les interactions entre les personnages et les enceintes connectées dans les films publicitaires, nous remarquons l'apparition récurrente de certains mots. Dans l'ensemble des publicités étudiées, les protagonistes utilisent un *smart speaker* pour lui demander, par exemple, d'allumer la lumière, mettre de la musique ou répondre à une question. Les enceintes connectées, répondant à une série limitée d'ordres, sont programmées pour associer certains mots à certaines actions. Dès lors, imprégner l'utilisateur du vocabulaire compris par le logiciel devient un impératif pour une bonne utilisation. Les films, qu'ils soient de Google, Amazon ou Apple, mettent en scène les mêmes actions et y associent les mêmes mots. Les deux exemples les plus percutants sont « shopping list » et « playlist ». Nous retrouvons un de ces mots dans la totalité des publicités analysées. Selon la définition du Larousse, une *playlist* est une « liste des disques diffusés, à un moment donné, par une station de radio » et une *shopping list* est définie comme « a list of the things that you want to buy when you go shopping, which you write on a piece of paper », selon le site du dictionnaire Collins. Ces deux termes propres à la consommation en ligne correspondent à l'objectif des enceintes connectées : l'achat. Lors de notre entretien avec Yann Lechelle, COO de Snips, celui-ci affirmait que les utilisateurs savaient qu'ils pouvaient demander la météo à leur enceinte, uniquement parce qu'ils avaient vu la publicité à la télévision qui montrait cette fonction. Les GAFAs ont conscience de l'impact des publicités et profitent ainsi de cette communication pour mettre en avant les fonctions génératrices de consommation. Si l'on prend l'exemple de la première fonction de l'enceinte, la musique, il n'est possible de mettre de la musique sur un HomePod que si son possesseur a un compte Apple Music. Google n'accepte de son côté, que la musique provenant de Google Play Music, Spotify ou Deezer et Amazon Echo ne lit que la musique d'Amazon Prime Music ou Spotify. Ce fonctionnement qui peut paraître restrictif correspond à la stratégie de « verrouillage » des clients à qui les GAFAs mettent « des barrières à la mobilité ». Les GAFAs incitent les consommateurs à toujours utiliser la même

marque pour devenir le standard technologique. Ils mettent en place une stratégie de *lock-in*, c'est-à-dire qu'ils bloquent leur concurrent<sup>36</sup>. Ainsi, avec un iPhone, on pourra facilement retrouver sa musique sur son ordinateur en achetant un Mac, et sur son enceinte connectée si l'on a un HomePod. Cela fonctionne pour la musique, mais aussi pour toutes les applications liées entre les différents appareils : les mails, le navigateur web, les notes, le calendrier, et plus généralement, le cloud. Une entreprise peut mettre en place une stratégie de *lock-in* lorsqu'elle détient une majorité du marché et peut alors faire pression sur ses concurrents en leur imposant l'utilisation de son produit.

Les récurrences mises en avant dans les films publicitaires ne sont donc pas des cas fortuits mais placés dans l'objectif de communiquer aux consommateurs ce qu'ils doivent et comment ils doivent consommer.

---

<sup>36</sup> Bouvier-Patron Paul, « L'application des concepts de 'lock-in' et de 'barrières à la mobilité' à une théorie des réseaux d'entreprises », *Revue française d'économie*, volume 9, n°1, 1994. pp. 205-232.

## B. Familiariser aux fonctionnalités

Les enceintes connectées émergent dans un marché porteur. En 2015, Google annonçait que les requêtes sur mobile avaient dépassé celles sur ordinateur aux États-Unis et au Japon, et dans 8 autres pays, et depuis ce chiffre ne fait qu'augmenter. Avec 1,472 milliard de smartphones vendus en 2017, ce marché se porte bien. En intégrant des assistants vocaux aux téléphones, les recherches par commande vocale se sont progressivement installées. Lors d'un entretien chez Google, Sébastien Fourault, « mobile & Assistant specialist » nous explique qu'en France, 64% des utilisateurs connaissent Google Home et que 32% l'ont déjà utilisé. Selon leurs études, les usages récurrents sont la musique, la domotique, la lecture d'informations, le minuteur et la liste de courses, puis « les usages fun », comme les blagues. Autant de services qui sont montrés dans les publicités pour enceinte connectée. Pourtant, lors de la dernière conférence Google I/O, Sundar Pichai, CEO de Google, affichait fièrement que 500 millions d'objets étaient connectés au Google Assistant, incluant les voitures de 40 marques différentes. Amazon de son côté compte 40 000 applications vocales, les *skills*, dont Marmiton, annuaire de recettes de cuisine, Air France et Ouisncf, ou Foodchéri, un système de livraison de repas à domicile. Le groupe a également lancé un outil<sup>37</sup> permettant aux utilisateurs de ses enceintes Echo de créer leurs propres *skills* Alexa, et ce sans nécessiter de connaissances en développement. Il s'agit d'une bibliothèque libre d'accès dans laquelle sont déjà stockés plusieurs modèles de *skills* personnalisables.

Steve Rabuchin, le vice-président d'Amazon a déclaré dans un communiqué de presse : « Vous n'avez pas besoin d'acquérir de l'expérience ou de savoir coder pour commencer. Ma famille a créé notre propre blague et ça a été génial d'interagir avec Alexa d'une manière totalement nouvelle et personnelle ». Les différentes possibilités sont classées ainsi : *Fun & Games*, *Learning & Knowledge*, *At home*, et *Storyteller*. Vous pouvez créer par exemple, une petite histoire, un guide pour la *baby-sitter* ou un quiz. Les assistants vocaux permettent ainsi un grand périmètre d'actions qui risque de s'étendre dans les prochaines années, mais les utilisations réelles semblent pourtant éloignées de toutes ces possibilités techniques. Comment les fabricants communiquent-ils sur les différentes fonctions des smart speakers ? L'utilisation limitée des consommateurs est-elle la preuve d'une mauvaise communication ?

Nous verrons dans cette partie dans quelle mesure les enceintes connectées se confrontent au problème de l'affordance et du signifiant, puis nous analyserons les stratégies qu'ont adoptées les marques pour contourner le problème. Pour comprendre et utiliser les notions

---

<sup>37</sup> Capture d'écran du site web en annexe 2

d'affordance et de signifiant, nous baserons notre étude sur le texte de Donald Norman, *The Design of Everyday Things*<sup>38</sup>.

### 1. Le problème de l'affordance et des signifiants

Dans son ouvrage *The Design of Everyday Things*, Donald Norman nous parle d'interaction avec nos objets du quotidien. L'auteur part du fait que nous rencontrons tous les jours de nouveaux objets, et pourtant, pour la plupart, nous « devinons » comment ils fonctionnent. Nous apprenons à nous servir de ces objets à travers ce qu'il appelle la *discoverability*, traduit par la découvrabilité. La découvrabilité est le résultat de cinq concepts psychologiques : l'affordance, le signifiant, la contrainte, le mapping et le feedback.

L'affordance est définie comme « the relationship between a physical object and a person (or for that matter, any interacting agent, whether animal or human, or even machines and robots). An affordance is a relationship between the properties of an object and the capabilities of the agent that determine just how the object could possibly be used <sup>39</sup> ». L'affordance d'un objet est sa capacité à communiquer son utilité. Une porte, sans poignée avec une plaque en métal, insinue qu'il faut la pousser. Une fente incite à glisser quelque chose à l'intérieur. Une balle semble faite pour rebondir, bouger, être balancée. Toutes ces affordances visibles permettent à l'utilisateur de comprendre le fonctionnement d'un objet sans instruction. La réflexion de Donald Norman est à l'origine des études de J.J Gibson, le psychologue qui a inventé le terme d'affordance. Celui-ci part du principe que tous les sens travaillent ensemble : la vue, le son, l'odeur, le toucher, l'équilibre, la kinesthésie et la position de notre corps. Tous ces sens ramassent une quantité d'informations et ainsi déterminent nos perceptions sans besoin d'un traitement interne ou de cognition, sans faire appel à la réflexion. Donald Norman apporte un complément aux études du psychologue et insiste sur l'aspect relationnel de l'affordance : celle-ci existe en fonction des propriétés de l'objet mais aussi de l'utilisateur. Donald Norman prend l'exemple du verre qui est une structure qui bloque le passage et permet de voir à travers. On peut alors parler d'*anti-affordance*, une contradiction dans l'objet même. Un objet, pour être utile, doit laisser percevoir son affordance comme son anti-affordance. Ainsi, si l'on prend le cas d'une baie vitrée, celle-ci devrait être transparente tout en signalant sa présence pour éviter aux utilisateurs de se cogner dedans. L'affiche « Pull » qui pourrait être affichée sur une porte

---

<sup>38</sup> Donald Norman, « The Design of Everyday Things », 1988, New-York, Basic Books

<sup>39</sup> Donald Norman, « The Design of Everyday Things », 1988, New-York, Basic Books – Chapter One, The Psychopathology of Everyday Things, page 11

baie vitrée deviendrait ce que l'on appelle un *signifiant*<sup>40</sup>. Un signifiant spécifie comment l'utilisateur découvre les différentes fonctionnalités : ce sont des signes perceptibles de ce qui peut être fait. Sur une interface graphique par exemple, un bouton « démarrer » n'est pas de l'affordance mais un signifiant. L'affordance détermine les actions possibles alors qu'un signifiant indique où l'action se situe précisément. Pour un design complet, les deux sont essentiels. Le signifiant peut être intentionnel ou accidentel. L'auteur prend l'exemple d'un panneau « Push » sur une porte comme signifiant intentionnel et les marques de pas sur la neige qui indique un chemin comme signifiant accidentel. Peu importe que le signifiant soit intentionnel ou non, l'important est qu'il existe.

Si nous revenons au cas des enceintes connectées, leur design guident-ils l'utilisateur ?

a. Analyse du design des enceintes connectées : étude de cas avec Google Home

En analysant la notice du Google Home, nous remarquons que celle-ci ne donne que très peu d'éléments sur son fonctionnement :

*« Étape 1 : Branchez le câble d'alimentation à votre appareil Google Home*

*Étape 2 : Branchez l'adaptateur secteur sur une prise murale*

*Étape 3 : Configurez votre appareil Google Home*

*Téléchargez et exécutez l'application Google Home sur un téléphone ou sur une tablette.*

*Configurez votre appareil.*

*Étape 4 : Commencez à parler à l'Assistant Google sur Google Home ou Google*

*Home Mini.*<sup>41</sup> »

Une fois branchée, il « suffit » de parler à son Google Assistant. La marque considère ainsi que son produit est intuitif et qu'une notice d'utilisation n'est pas nécessaire. Est-ce par effet de mode, puisqu'une interface doit maintenant être comprise par tous pour être considérée comme efficace ? Analysons l'interface d'un Google Home et tentons de comprendre quels signifiants l'appareil offre à son utilisateur pour le guider.

---

<sup>40</sup> Dans le texte original anglais, le mot utilisé est « signifier »

<sup>41</sup> Notice prise sur le site de vente Google de Google Home



*Image de gauche : Google Home vue de dos - image de droite : Google Home vue de face*

Vue de face, le Google Home est lisse et n'offre aucun bouton. On comprend par la matière grise en tissu qu'il s'agit de la partie stéréo de l'assistant, là d'où sort le son. Cette partie, par son tissu et sa couleur propre aux enceintes musicales rappelle la fonction principale du produit : l'émission de son. Dès lors, le sous-bassement gris de l'enceinte crée une affordance dans la mesure où elle permet à l'utilisateur d'identifier à quoi elle pourrait probablement servir même si, sans électricité, rien ne le confirme. Ce sont ici les codes culturels qui permettent à l'utilisateur de reconnaître la matière et la couleur, propre aux enceintes classiques, et ainsi de décoder cette première fonction. La partie blanche reste neutre et semble agir comme un socle inversé. La face oblique qui affiche des ronds de couleur lorsque l'enceinte se met en route ou lance une action intervient ici comme un écran. Son orientation donne à l'utilisateur la possibilité de vérifier visuellement si l'enceinte a bien démarré ou a bien pris en compte sa commande. L'apparition des voyants lumineux lorsque l'on prononce « Ok Google » ou lorsque l'on s'adresse à l'appareil s'appelle le feedback. Le feedback communique le résultat d'une action. Le feedback apparaît pour la première fois en communication dans les travaux de Norbert Wiener, mathématicien fondateur de la cybernétique. Le scientifique définit la cybernétique comme « la théorie entière de la commande et de la communication, aussi bien chez l'animal que dans la machine<sup>42</sup> ». Il considère que le monde est constitué de systèmes, vivants ou non-vivants, imbriqués en interaction. Ces systèmes communiquent selon le schéma suivant :

*Source émetteur → Canal → récepteur*

---

<sup>42</sup> Norbert Wiener, *La cybernétique : Information et régulation dans le vivant et la machine*, Seuil, 2014, « Introduction », p. 70

Lorsque le récepteur reçoit le message celui-ci renvoie un feedback qui conduit soit à accentuer le phénomène de communication soit à le réguler, à l'amortir. Selon Donald Norman, le feedback est indispensable dans l'interaction avec un objet. Celui-ci doit être immédiat pour ne pas décourager l'utilisateur à comprendre le produit. Google utilise donc ici un feedback visuel. Quand on dit « Ok Google », l'appareil fait apparaître des petites leds de couleur qui permettent la vérification du bon fonctionnement pour l'utilisateur. Le signal visuel a un double rôle en matière de feedback puisque c'est également sa désactivation qui permet à l'utilisateur de savoir quand la conversation est terminée. Ainsi, au-delà de l'interaction avec la voix, l'interaction visuelle reste nécessaire pour renvoyer un maximum de feedback et rassurer l'utilisateur habitué à l'interface graphique.

Sur la face arrière, Google a introduit un bouton *mute* qui permet de désactiver le micro lorsque l'utilisateur ne souhaite pas être écouté. Ce bouton est un signifiant, premièrement parce que c'est un bouton, sa forme creuse permet de percevoir qu'il peut être enfoncé et le pictogramme « micro coupé » indique son utilité. Ainsi, si le sous bassement gris de l'enceinte permet par les codes culturels de classer l'objet comme émetteur de son, le bouton *mute* indique sa deuxième fonction, l'enregistrement de son.

Physiquement, ce bouton *mute* est le seul signifiant de l'enceinte mais il n'est pas la seule fonction s'activant par le toucher. En effet, la face oblique est tactile et permet en la tapotant d'arrêter l'action en cours (de la musique par exemple) ou d'augmenter et baisser le volume de l'appareil. Une seconde fois, l'enceinte connectée ne limite pas son utilisation à la voix et au son et stimule un autre sens en proposant une utilisation par le toucher.

Le design de l'enceinte connectée permet de créer une relation d'affordance dans la mesure où il permet de comprendre les deux fonctions principales de l'appareil : l'émission et l'enregistrement de son. Son écran tactile affichant des ronds lumineux permet un feedback essentiel à son fonctionnement. L'objet en lui-même remplit donc son rôle. En revanche, rien ne nous indique réellement ce que peut faire la machine, le logiciel Google Assistant. C'est ici que se pose la question de l'affordance : comment créer une interaction entre un logiciel et son utilisateur sans interface graphique ?

## 2. La stratégie des marques

Lors de notre entretien avec Yann Lechelle, COO de Snips, nous avons évoqué le problème de l'affordance que rencontrent les enceintes connectées, support d'un assistant vocal sans interface graphique. Il existerait un décalage entre ce que nous projetons d'un assistant connecté et la réalité. Il est vrai que l'industrie culturelle, à commencer par les films de science-fiction, ont propagé une certaine image de l'intelligence artificielle et particulièrement de celle des robots parlant. De HAL 9000, l'ordinateur parlant de *2001, l'Odyssée de l'espace* à Samantha, l'intelligence artificielle du film *Her*, les machines parlantes nous apparaissent comme des « similis humains<sup>43</sup> » avec qui il est possible d'avoir une conversation, voire une relation<sup>44</sup>. En vendant une enceinte connectée, les marques doivent se confronter à l'imaginaire commun et anticiper les attentes des consommateurs persuadés qu'un Google Home pourra devenir leur *réel* assistant personnel.

*« Siri, c'est finalement de dire que c'est l'assistant d'un homme, à qui on peut poser toutes les questions et ça c'est un problème d'affordance. [...] un assistant vocal, peu importe la forme, c'est la boîte de pandore. On se dit "je peux tout lui demander". »*

*Extrait de l'entretien avec Yann Lechelle, COO de Snips*

Nous imaginons ainsi que nous pouvons tout demander à notre assistant, qu'il s'agisse d'un simple ordre, un conseil ou même un compliment. Pourtant, la réalité est encore éloignée de l'idée que nous nous faisons d'une conversation avec un robot. Pour l'instant, si nous souhaitons interagir avec une enceinte connectée, nous devons l'activer en prononçant un *wake-word* et, comme nous l'avons vu précédemment, adapter notre langage à la machine pour être certain que celle-ci comprendra notre requête. Pour éviter une trop grande déception du consommateur, les fabricants ont propagé une communication pédagogique visant à montrer ce qui est possible ou pas avec une enceinte connectée.

Ainsi, quels moyens communicationnels les marques ont-elles mis en place pour lancer un produit fantasmé depuis plusieurs générations ?

---

<sup>43</sup> Voir annexe 1, interview de Yann Lechelle

<sup>44</sup> Dans le film *Her* (2013) de Spike Jonze, le personnage principal tombe amoureux de son intelligence artificielle

- a. Les assistants vocaux sur smartphone, une aide précieuse dans l'éducation aux enceintes connectées

Quatre ans après la sortie de Siri, 65% des utilisateurs de smartphones avaient déjà utilisé un assistant vocal aux Etats-Unis<sup>45</sup>. La voix, à travers l'émergence des assistants vocaux sur smartphone, s'est ainsi progressivement développée et tend aujourd'hui à croître. Selon ComScore, entreprise américaine d'analyse publicitaire, 50% de nos recherches se feront par la voix d'ici 2020. En 2018, la société d'audit et de conseils PriceWaterhouseCoopers (PwC), a interrogé un échantillon de 1000 Américains âgés de 18 à 64 ans et qui ont accès à Internet et a organisé deux groupes de discussion pour échanger au sujet de l'utilisation des assistants vocaux. Leur étude montre premièrement que les assistants vocaux et les supports liés sont majoritairement connus parmi les sondés. Les utilisateurs des assistants vocaux de smartphone sont donc les mêmes que les utilisateurs d'enceinte connectée : les ménages entre 25 et 49 ans. Ces usagers font aujourd'hui 57% de leurs requêtes vocales sur smartphone contre 29% sur tablette et ordinateur et 27% sur enceinte connectée. L'éducation à l'interface vocale est donc premièrement assurée par les fabricants de smartphone, en particulier par Apple avec Siri. L'intégration d'une interface vocale à une interface graphique, connue des utilisateurs, a permis d'intégrer cette pratique dans le quotidien des consommateurs. Dans son mémoire, Clotilde Chevet évoque l'écriture de la parole de l'utilisateur :

*« A l'inverse de la parole de l'entreprise, la parole de l'utilisateur est spontanée, elle est immédiate et orale avant d'être écrite. Pourtant on se rend très vite compte que la parole de l'utilisateur est contrainte par l'interlocuteur qu'est la machine. Elle doit être formulée dans un style écrit de façon à être comprise par la machine. »*

*Extrait du mémoire de Clotilde Chevet, page 75*

La retranscription écrite des paroles de l'utilisateur à l'écran lui a ainsi permis de comprendre le fonctionnement de la machine et de s'adapter à ses codes : des questions ou des ordres clairs, une diction parfaite, pas de marque d'hésitation ou de correction de sa propre phrase. L'utilisateur a saisi que pour être audible, sa parole devait pouvoir être correctement retranscrite à l'écrit. Afin de créer l'illusion d'une conversation, la machine humanise sa réponse en utilisant une rhétorique orale<sup>46</sup> et répond ainsi à sa promesse d'un

---

<sup>45</sup> Pestanes Philippe, Gauthier Benjamin, « Essor des assistants intelligents : nouveau gadget pour votre salon ou fenêtre d'opportunité pour rebattre les cartes de l'économie du web ? », 2017, étude réalisée par le cabinet Wavestone

<sup>46</sup> Clotilde Chevet, "La part de l'imaginaire dans « l'interaction homme-machine » : le cas des assistants personnels numériques", Master 2 Recherche Mention Information et Communication, sous la direction du Pr Karine Berthelot-Guiet, Paris, Université de Paris IV-Sorbonne - CELSA, 2015.

dialogue fluide et authentique. Au-delà de la transmission de la technique, les assistants vocaux sur smartphone ont également permis à sensibiliser les consommateurs aux possibilités d'usage d'un assistant vocal. Malgré une forte évolution de la compréhension du langage et l'émergence de nouvelles fonctionnalités des assistants vocaux, les fantasmes de l'assistant intelligent avec qui l'on peut converser comme avec un ami se sont peu à peu atténués. Pourtant, si le logiciel interne reste le même, les marques semblent s'investir davantage dans l'éducation aux enceintes connectées pour renforcer cette pratique et créer de réelles habitudes de consommation. Comment tentent-elles d'influencer notre usage pour espérer relocaliser une partie de nos actions numériques sur l'interface vocale ?

b. Une concentration de la communication sur une série limitée d'actions

En étudiant les publicités et discours qui accompagnent la sortie des enceintes connectées, nous remarquons que certaines fonctions sont beaucoup plus représentées que d'autres. Pour arriver à ce résultat, nous avons construit un tableau et noté pour chaque film publicitaire (publicités Google ou Amazon, de 2015 à 2018) ou article, les fonctions attribuées à l'enceinte. Les fonctions mises en scène dans ces contenus sont : la musique, la météo, l'appel téléphonique, les notes et les rappels, la « recherche Google », la liste de courses, la domotique, le calendrier, l'agenda, le minuteur, la blague et le GPS. Selon notre analyse, c'est la « recherche Internet » qui remporte le classement, cette fonction est principalement mise en avant, que ce soit dans les publicités des marques ou dans les articles de presse consacrés aux enceintes connectées. La musique arrive en seconde position et la domotique en troisième position. Ce classement correspond à une étude menée par le cabinet Capgemini<sup>47</sup> sur l'usage des assistants vocaux. En effet, leur étude montre que 82% de l'usage des assistants vocaux (attention, ici nous parlons sur tous supports confondus) est la recherche d'informations, 67% pour mettre de la musique ou lancer une vidéo, 36% pour accéder à un service d'une marque et 31% pour la domotique<sup>48</sup>.

Il existe ainsi une corrélation entre la communication des marques et l'utilisation des consommateurs. Mais pourquoi les fabricants communiquent-ils majoritairement sur la recherche d'information et la musique ?

---

<sup>47</sup> Extrait de l'étude en annexe 3

<sup>48</sup> Capgemini Digital Transformation Institute, Conversational Commerce Survey, October–November 2017, N = 2,558 users in the US, UK, France, and Germany.

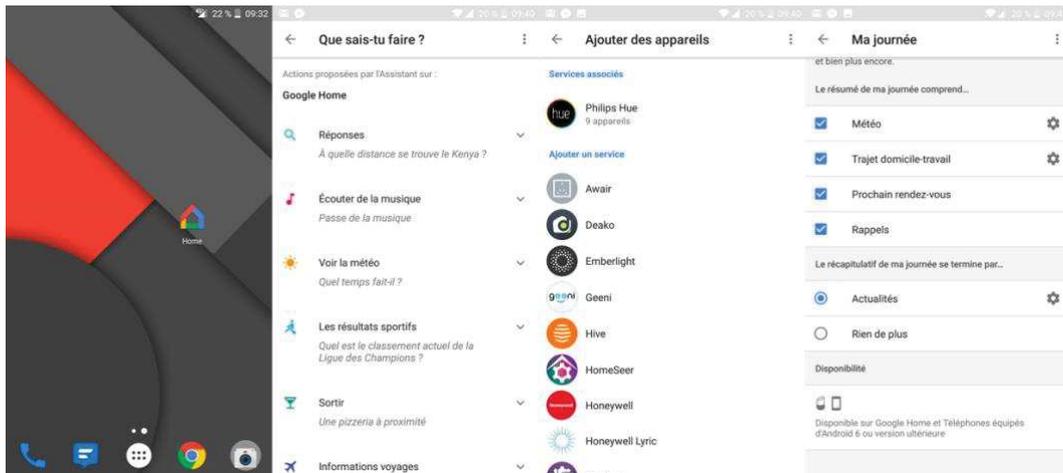
Comme nous l'avons vu précédemment, Google annonçait fièrement lors de sa dernière keynote que le Google Assistant pouvait se connecter à 500 millions d'objets dont les voitures de 40 marques différentes. La domotique semble être le défi des fabricants pour réussir à pénétrer dans le quotidien des consommateurs. Google vend même des packs de domotique, comme un Google Home et une ampoule connectée Phillips. Pourtant, la communication des marques ne suit pas encore clairement cette directive mais insiste plutôt sur l'utilisation de l'assistant personnel pour toutes sortes de recherche Internet : actualités, météo, recettes, rendez-vous (relié à un agenda en ligne), offres commerciales et même « conseils relationnels » - toujours selon l'étude de Capgemini. Les marques investissent ainsi davantage sur les actions déjà connues des utilisateurs pour intégrer la pratique dans leur quotidien et les initier à l'usage de la commande vocale.

Les publicités mettent aussi en avant cette fonction pour valoriser l'atout principal de l'interface vocale : pouvoir faire deux choses en même temps. En effet, les marques mettent régulièrement en scène une femme ou un homme entrain de cuisiner ou de s'occuper de ses enfants, des moments où les mains sont donc occupées. Ainsi, les marques concentrent leur communication sur trois fonctions principales pour transmettre aux utilisateurs ce qu'ils peuvent faire avec leurs enceintes et à quel moment celles-ci peuvent leur être utiles.

### **C. Habituer à l'absence d'interface graphique**

Si notre rapport à l'informatique s'est majoritairement développé avec les interfaces graphiques, c'est parce que celles-ci ont connu un développement important et ont su s'adapter aux consommateurs en créant la souris et le clavier, lien entre la machine et l'utilisateur. Les souris des premiers ordinateurs, le Xerox Alto et le Star étaient respectivement munies de trois et deux boutons. La souris de l'ordinateur Lisa d'Apple n'en présentait qu'un seul, ce qui a entraîné l'apparition du double clic et du « drag & drop » (sélectionner, glisser), permettant ainsi de faire différentes actions avec un seul bouton. Apple a également modifié et simplifié les interfaces en les rendant plus accessibles, plus intuitives. Lisa a par exemple introduit la coche qui permet de voir quand une action est en route. Microsoft, autre acteur du marché de l'informatique a également révolutionné les interfaces graphiques en introduisant les couleurs et en permettant à deux fenêtres de se superposer. Si

l'interface vocale était également en développement dans ces entreprises, c'est l'interface graphique qui a primé jusqu'aujourd'hui encore. Les fabricants comme Apple ou Google ont d'ailleurs premièrement intégré leurs assistants vocaux en les greffant à une interface graphique sur smartphone. Aujourd'hui, l'utilisation d'une enceinte connectée passe encore par une interface graphique dans la mesure où les paramètres se règlent sur une application dédiée.



Captures d'écran de l'application Google Home<sup>49</sup>

Une fois l'enceinte paramétrée, l'objectif est d'utiliser sa voix et non plus le toucher même s'il est encore possible de changer de musique via son smartphone ou d'éteindre la lumière avec un interrupteur, des gestes ancrés dans nos quotidiens. Alors, comment bouleverser les habitudes des consommateurs et les convaincre d'oublier ce qu'ils faisaient si naturellement pour tout contrôler par la voix ?

#### 1. Chez Amazon : la mise en valeur d'une distance entre utilisateur et technologie

En observant le dispositif de communication d'Amazon, on remarque que c'est la simplicité qui est donnée comme premier argument de vente. En 2015, un an après la sortie du produit, le slogan publicitaire de la firme était « Just ask », en français cela pourrait se traduire par « Demandez, tout simplement ». Le slogan rappelle qu'avec l'enceinte connectée, il suffit de poser une question ou donner un ordre sans avoir à sortir son portable, une feuille de papier ou son agenda pour noter, mettre une alarme, ou obtenir une information. En réduisant le nombre d'actions, la marque sous-entend que chaque petit geste du quotidien devient plus rapide, plus fluide et moins prenant. Dans son film publicitaire sorti en juin 2015,

<sup>49</sup> Plus d'illustration en annexe 2

la marque réalise un reportage fictif chez des utilisateurs d'Alexa. Filmés dans leur vie quotidienne, ils vantent les fonctions d'Alexa. L'enceinte connectée est ici implicitement comparée à l'interface graphique. La marque dénonce une dépendance à l'écran qui complique les tâches quotidiennes plus qu'elle ne nous aide. Le premier argument repose sur la dépendance matérielle, le fait de devoir toujours avoir son portable, ordinateur ou tablette sur soi ne facilite pas un usage fluide. Le film met en scène un homme âgé qui se réjouit de pouvoir utiliser une enceinte connectée lorsqu'il est devant sa télévision et n'a pas le courage de se relever. La voix off explique au 'Alexa nous écoute peu importe où nous sommes et est donc toujours prête à nous aider, une vraie plus-value par rapport aux smartphones, ordinateurs et tablettes. Un autre argument pour l'interface vocale est l'argument physique. Le film met en scène un couple de personnes aveugles qui confient qu'Alexa est devenue nécessaire à leur quotidien, pour le minuteur par exemple. Amazon vend la voix comme une nouvelle pratique qui rend la technologie plus accessible et plus intégrée, presque invisible. Le projet de rendre les interfaces invisibles a débuté au Palo Alto Research Center, en Californie dans les années 80<sup>50</sup>. La vision d'Amazon qui prétend à effacer l'interface correspond à l'idée de l'informatique personnel, concept né sous l'impulsion de la contre-culture américaine des années 80. L'informatique personnel remet en cause la vision d'un informatique confinée dans un objet technique. Nous pourrions rapprocher ce concept des paradigmes d'Alan Cooper et ainsi considérer que l'informatique personnel des années 80 va à l'encontre d'un paradigme technique qui met une barrière entre la machine et son utilisateur. Les ingénieurs américains des années 80 cherchaient à faire de l'utilisateur celui qui déterminerait l'usage qui sera fait de la machine. En d'autres termes, cette idéologie rendrait à l'utilisateur la liberté de définir comment utiliser un ordinateur et non l'inverse. Cette pensée met l'informatique au cœur de nos quotidiens pour faire en sorte que celui-ci devienne augmenté et au plus près de notre intimité.

« Nous aurons réussi lorsque le monde sera devenu notre propre interface »

*Neil Gershenfeld, directeur du Things that Think<sup>51</sup> (1999)*

La vision d'Amazon s'inscrit dans ce courant de pensée dans la mesure où la marque place son produit comme un outil invisible et nécessaire. Alexa semble accessible à tous et

---

<sup>50</sup> David Ducheu, « Effacer l'interface. Une trajectoire du design de l'interaction homme-machine » Interfaces numériques, volume 5 n°2/ juin-septembre 2016, Designers interactifs (page 257 - 276)

<sup>51</sup> Think tank du MIT

révolutionnaire en facilitant chaque petit geste : la recherche Internet, le minuteur, le calendrier, le GPS, la to-do list et la musique. L'écran est banni de l'environnement du film. Il n'intervient que pour montrer que les modifications demandées à Alexa se synchronisent correctement avec un smartphone. Le manque créé par l'absence d'interface graphique est ainsi résolu en communiquant sur l'idéologie d'une technologie parfaitement intégrée à notre quotidien.

## 2. Chez Google : la promesse d'une reconnexion

### a. « Google Home – Profitez du moment »

Google présente l'interface vocale d'une manière différente de celle d'Amazon. En septembre 2017, l'entreprise américaine sort une série de publicités qui placent l'enceinte connectée comme moyen de déconnexion. La campagne publicitaire est nommée « Google Home – Profitez du moment » et positionne la déconnexion au cœur des préoccupations de Google. À travers ce titre, Google sous-entend que l'interface graphique captive son utilisateur, l'empêchant de communiquer avec le reste de son environnement. Dans une première vidéo, un jeune père, assis sur son canapé, navigue sur sa tablette. À côté, ses deux petites filles jouent. La voix off commente : « La maison est un endroit fait de moments uniques, comme celui-ci, là, juste là... et vous l'avez raté ». Cette remarque accuse le père de négliger les moments en famille pour préférer sa tablette. Google Home se positionne comme une solution au problème en supprimant l'interface graphique et en libérant ainsi les mains et le regard de l'écran. Lorsque le père, ébahi, réalise qu'il loupe des moments précieux, il quitte sa tablette pour aller jouer avec ses filles. Google Home entre alors en scène :

*« Si seulement quelque chose pouvait **nous aider à être plus présent** ?  
Hey, salut Google Home ! »*

L'enceinte permet au père d'effectuer une recherche sur Internet tout en profitant d'un instant avec ses enfants : « Ok Google, quelle taille fait la Tour Eiffel ? », « Ok Google, lance la playlist Valentine et Chloé ». En supprimant le rapport exclusif entre le père et l'écran, celui-ci retrouve un moment de complicité avec ses enfants. Google met ici en scène un homme de 30 ans environ. Il a grandi avec le développement d'internet et est habitué à avoir devant lui un écran d'ordinateur, de tablette ou de smartphone. L'enjeu du fabricant est donc

de bouleverser l'habitude du consommateur et de changer son rapport à la technologie en le faisant passer d'une relation très personnelle à ses objets numériques à un usage vocal qui ne permet pas d'intimité. L'ensemble des vidéos de la campagne « Profitez du moment » mettent en scène de jeunes adultes avec leurs enfants. La cible de Google sont les nouveaux parents, déjà connectés et habitués aux services des GAFAs. Au goûter d'anniversaire, ou pendant une séance de coloriage, les parents imaginés par Google participent aux activités, encouragent leurs progénitures et sont proche d'elles. L'objectif des fabricants est double. Dans un premier temps, il s'agit d'habituer les utilisateurs déjà consommateurs de smartphone et tablette à l'absence d'écran et leur faire comprendre que l'interface vocale correspond davantage à ce qu'ils recherchent : une *détox* des outils numériques, qui, nous le verrons, sont souvent associés au stress et au monde professionnel. Une fois l'enceinte connectée intégrée au foyer, le second objectif des fabricants est de se faire connaître à une population juvénile, les enfants – espérant créer avec eux une relation au statut de minorité, afin de rendre le savoir-faire technique de l'enceinte connectée implicite et non plus réfléchi.

b. L'émergence d'un discours alarmiste sur l'addiction aux écrans

Afin que l'usage devienne implicite et non réfléchi, la communication de Google répond à la tendance de la « digital detox ».

Le psychiatre Serge Tisseron étudie de près nos relations aux nouvelles technologies et leurs effets sur les enfants. Il publie de nombreux ouvrages sur la question, comme « Virtuel mon amour » et « Qui a peur des jeux vidéos ? » publiés en 2008, ou autre exemple, « Le jour où mon robot m'aimera. Vers l'empathie artificielle » en 2015. Serge Tisseron écrit régulièrement des articles pour la presse généraliste et participe à des émissions de radio sur France Inter par exemple. Si l'auteur n'est pas à l'origine du débat, sa présence dans les médias généralistes est une des preuves que les effets des nouvelles technologies intéressent le grand public. On assiste d'ailleurs de plus en plus à la naissance de stage de « digital detox », de bar sans wifi et, depuis le 1<sup>er</sup> janvier 2017 en France, l'article L2242-8 du Code du travail autorise un droit à la déconnexion – soit l'autorisation donnée à un salarié de ne pas répondre à ses mails ou à son portable après une certaine heure. Serge Tisseron écrit pourtant en 2014, « Les programmes intensifs de detox numérique permettront sans doute aux uns et aux autres de penser à d'autres choses pendant une semaine ou deux. Mais que se passera-t-il à leur retour ? Toutes les habitudes associées à leur environnement habituel surgiront à

nouveau et la tentation sera grande de renouer avec elles »<sup>52</sup>. Le psychiatre confirme qu'il est bien difficile de se séparer de son smartphone, l'objet étant à présent imprégné dans nos quotidiens. Vérifier ses notifications et scroller à l'infini seraient presque devenus des gestes infra-ordinaires.

Néanmoins, ces gestes numériques se sont révélés être de mauvaises habitudes : chez les adolescents et adultes, l'usage abusif de smartphone serait lié à une augmentation de l'anxiété et de la dépression. Les chercheurs Erik Peper et Richard Harvey de l'ISNR (International Society for Neurofeedback & Research) ont publié en mars 2018, une étude sur notre addiction aux smartphones. Les scientifiques prouvent que ces appareils monopolisent bien notre attention et dès lors, nous déconcentrent et nous isoleraient. Selon l'étude, sur 135 étudiants interrogés, ceux qui utilisent le plus leur téléphone enregistrent les plus importants degrés d'anxiété. Les chercheurs vont même jusqu'à assimiler l'addiction au smartphone à celle de l'opium<sup>53</sup>. Ils insistent sur l'effet des notifications et de ce qu'elles déclenchent chez l'utilisateur : une stimulation neurologique, comme lorsque l'on se fait prendre par surprise ou que l'on signale un danger dans notre environnement. Erik Peper et Richard Harvey citent l'article des chercheurs Jemin Lee, Jinse Kwon et Hyungshin Kim de l'Université de Chungnam, en Corée du Sud. Les auteurs de « Reducing distraction of smartwatch users with deep learning » soulignent l'aspect contradictoire de la notification : nous devons à la fois y répondre pour exister socialement mais éviter d'en recevoir un nombre démesuré pour continuer aussi à exister socialement :

*« [...] so the balance between a constant demand to orient towards a notification versus demand to orient towards a notification versus ignoring all notifications requires choices by the users of smartphone technology. »<sup>54</sup>*

### c. La déconnexion, le nouveau défi de la Silicon Valley

Les publicités pour enceinte connectée sont des exemples d'une prise de position face à l'addiction des écrans mais ce discours se vérifie aussi par la mise en place de dispositifs de « déconnexion » des GAFAs.

Dans la conférence Google I/O 2018, la conférence annuelle de Google pendant laquelle l'entreprise présente ses dernières innovations, le CEO Sundar Pichai évoque, à la

---

<sup>52</sup> Article du blog de Serge Tisseron : <http://sergetisseron.com/blog/detox-numerique-halte-a-l-intox/>

<sup>53</sup> Peper Erik, Harvey Richard, « Digital addiction : Increased loneliness, anxiety and depression », *NeuroRegulation*, n°5, pp 3-8

<sup>54</sup> Lee, Kwon et Kim repris par Peper Erik, Harvey Richard, « Digital addiction : Increased loneliness, anxiety and depression », *NeuroRegulation*, n°5, pp 3-8

suite d'une longue présentation sur Google Assistant et les objets connectés, le FOMO – Fear of Missing Out. Sundar Pichai parle alors de « digital detox » en déclarant qu'« il y a une pression croissante [qui pousse] à répondre immédiatement » et que « nous avons l'occasion de faire mieux » et de passer de « la peur de manquer quelque chose à la joie de manquer quelque chose », le JOMO - Joy of Missing Out. Pour lui, il y a quatre façons d'aider à développer le *digital wellbeing* : comprendre les habitudes des utilisateurs, se concentrer « sur ce qui compte vraiment », désactiver les notifications et permettre ainsi de trouver un équilibre avec sa famille :

*Extrait du slide de la conférence Google I/O 2018*



Pour promouvoir « le bien-être numérique », Google lance le *Dashboard*<sup>55</sup>, un tableau de bord intégré au nouvel Android nommé Android P. Il s'agit d'une série de commandes qui permettent d'indiquer le temps passé sur les applications et le nombre de notifications envoyées quotidiennement par chacune d'entre elles pour aider l'utilisateur à mesurer sa consommation. Il permet également de programmer des limites de temps d'utilisation, par exemple pas plus de 15 minutes par jour sur Facebook.

De son côté, Facebook a déclenché le même type de dispositif : chaque utilisateur peut connaître son temps d'activité sur ses différents appareils et choisir ou non de le limiter. Il est aussi possible d'installer un système d'alerte pour prévenir l'utilisateur qu'il dépasse son temps autorisé sur l'application ou encore de désactiver les notifications à partir d'une certaine heure. Pour éviter le *scrolling* infini, l'application prévient l'utilisateur pour lui dire qu'il a est rendu là où il s'était arrêté la fois précédente et évite ainsi de revoir les mêmes contenus. Technique de communication ou prise de conscience de la limite des écrans ? Peu importe puisque chez Google la solution semble toute trouvée : conforter le consommateur dans son souhait de déconnexion en l'orientant progressivement vers l'interface vocale.

---

<sup>55</sup> Voir annexe 2

## CONCLUSION DE LA PARTIE I

L'objectif de cette première partie était de déconstruire les discours didactiques qui accompagnent la commercialisation des enceintes connectées pour comprendre en quoi celles-ci nécessitent une éducation pratique. Notre première hypothèse reposait sur le fait que malgré l'essence naturelle de la parole, les enceintes connectées rencontrent un problème d'affordance et nécessitent ainsi une éducation sur leur simple fonctionnement.

Nous avons vu dans un premier temps les différents niveaux relationnels entre homme et technique à travers l'ouvrage de Gilbert Simondon, *De l'existence des objets techniques*. De par notre âge et notre culture, nous n'avons pas le même niveau de connaissance face à une machine. En étudiant les publicités des fabricants, nous avons perçu l'analyse marketing : la cible des enceintes connectées est un homme/une femme entre 25 et 45 ans, ayant un accès à Internet et surtout, utilisateurs de smartphone et tablette. Les communications des marques se veulent très pédagogiques. Nous en déduisons ainsi que les marques considèrent que les cibles sont dans une relation à statut de majorité avec les enceintes connectées. Les consommateurs doivent être éduqués parce que cet objet est encore inconnu et son utilisation nécessite une réflexion. Paradoxalement, la parole est censée être l'interface la plus naturelle puisqu'elle nous est enseignée dès notre plus jeune âge. Nous avons utilisé le concept d'*infra-ordinaire* de George Perec pour évoquer notre rapport au langage courant : un élément quotidien que nous ne remettons plus en question. Ces deux concepts mettent en évidence le paradoxe de la communication des marques d'enceintes connectées qui éduque les consommateurs adultes à un objet qui ne nécessite que l'usage de la parole. À partir de ce constat, nous avons analysé les dialogues entre les enceintes connectées et les personnages dans les publicités. Cette étude nous a permis de mettre en évidence le « langage de l'interface vocale », de la syntaxe au choix des mots, langage qui sert les intérêts économiques des marques. Dans la seconde partie, nous avons étudié les techniques de communication des marques sur les fonctions des enceintes connectées. L'assistance vocale étant en constante évolution, personne ne peut définir précisément à quoi une enceinte connectée peut servir. En se basant sur l'ouvrage de Donald Norman, *The Design of Everyday Things*, nous avons étudié le design de l'enceinte connectée Google Home et tenté de comprendre pourquoi celle-ci rencontrait un problème d'affordance. Dans un second temps, nous avons décrypté les techniques des marques pour répondre à ce problème qu'est la multiplicité de possibilité d'usages d'un assistant vocal. Nous avons vu que l'association de l'assistant vocal à un écran sur les smartphones avait d'abord permis à l'utilisateur de comprendre son fonctionnement et de déplacer

progressivement ses habitudes de l'écrit vers l'oral. Pour accentuer cette pratique, les marques ont insisté sur certaines fonctions dans leurs discours de communication : la recherche Internet, la musique et la domotique. L'étude de Capgemini<sup>56</sup> sur les usages des assistants vocaux, réalisée fin 2017, s'est penchée sur l'ensemble des supports comportant un assistant personnel. L'assistant vocal étant utilisé à 27% sur des enceintes connectées, nous considérons que l'étude est représentative de l'utilisation des enceintes connectées. On observe ainsi que la communication est en phase avec les usages. L'étude montre effectivement que les fonctions les plus utilisées sont la recherche Internet, la musique et vidéo et l'accès au service d'une marque. Pour ancrer la pratique de la commande vocale dans le quotidien des usagers, les marques ont repris ces habitudes et ont particulièrement concentré leur communication sur la pratique de la recherche Internet et la musique. La stratégie de communication des marques s'est ainsi révélée être un succès – les utilisateurs appliquent bien ce qu'ils connaissaient déjà et ce qu'ils ont vu à la télévision. Au-delà de transmettre un langage particulier et un certain nombre d'actions, les marques souhaitent également habituer les usagers à l'absence d'interface graphique. En revenant sur l'histoire de l'informatique et des interfaces, nous avons vu que l'interface graphique a été majoritairement présente dans les outils numériques grand public. Les premiers assistants vocaux, intégrés dans les smartphones, étaient, et sont toujours, associés à un écran et même si la possibilité d'interpeller son téléphone à distance est possible, des enquêtes montrent que cet usage est très peu fréquent. Chaque marque s'est positionnée pour prouver à travers sa communication que l'absence d'écran est une bonne chose pour chacun et permet plus de simplicité dans la gestion de son quotidien.

Comme nous l'avons vu en introduction, l'adoption des enceintes se fait rapidement – et leur stratégie d'éduquer par la publicité a probablement participé activement au phénomène. Pour autant, les marques ne cherchent pas à créer une adaptation dans un sens unique, de l'utilisateur vers la machine, mais elles cherchent à innover également en fonction des usages qu'elles observent. Dans sa dernière conférence annuelle, Google a annoncé vouloir supprimer partiellement le *wake word*. En effet, Scott Huffman, VP Engineering de Google, ironisait « « It gets a little annoying to say Hey Google everytime » en montrant la vidéo de la grand-mère « Goo-Goo »<sup>57</sup> dont on parlait précédemment. L'entreprise souhaite ainsi faire

---

<sup>56</sup> Capgemini Digital Transformation Institute, Conversational Commerce Survey, October–November 2017, N = 2,558 users in the US, UK, France, and Germany.

<sup>57</sup> Lien vers la vidéo : <https://www.youtube.com/watch?v=jEhRaPo2D1M>

une mise à jour pour permettre à l'utilisateur de ne dire qu'une seule fois « Ok Google » et poursuivre une conversation continue. Ce qui donnerait par exemple :

*Utilisateur - « Ok Google, quelle heure est-il ? »*

*Google Assistant – « Il est seize heures. »*

*Utilisateur – « Ajoute un rendez-vous avec le Docteur Pierre demain à onze heures. »*

*Google Assistant – « Ok, rendez-vous avec Docteur Pierre noté demain, mardi treize octobre à onze heures. »*

Dans la version actuelle, il faut dire « Ok Google » avant chaque demande, ce qui rend la conversation moins fluide et donc moins naturelle. Pour faciliter la conversation, Google souhaite également intégrer la fonction « multiple actions ». Cette mise à jour nous permettra de demander plusieurs choses à la fois au sein d'une seule phrase : « Ok Google, éteins la lumière du couloir et programme mon réveil pour sept heures demain matin. ». Aujourd'hui, nous devons faire deux demandes distinctes et prononcer deux fois le *wake word*.

Amazon de son côté lance des enceintes connectées avec un écran pour permettre une meilleure gestion de la domotique de la maison ou pouvoir lire une vidéo ou un texte si nécessaire. Les exemples montrés dans les premières vidéos de promotion sorties en février 2018 aux États-Unis montrent que l'écran peut par exemple être utilisé pour gérer les caméras de surveillance de la maison ou pour afficher une vidéo de cuisine.

On devine alors que les marques vont devoir adapter leurs discours et éduquer à nouveau sur l'évolution du « langage de l'assistance vocale » et ses pratiques.

## **II – Du discours bienveillant aux mises en garde étatiques : deux visions contradictoires sur la place des enceintes connectées dans les foyers**

Comme nous l'avons vu en première partie, l'assistant vocal est humanisé et personnalisé à travers le choix d'un certain langage. Qu'il soit une petite amie comme dans le film *Her* ou bien commandant de bord comme dans *2001, L'Odyssée de l'espace*, le robot parlant est fantasmé. L'absence d'un corps permet à l'utilisateur de l'imaginer de multiple façons et de lui donner différentes personnalités et donc différents rôles. Comme nous l'avons vu dans la précédente partie, la diversité des actions possibles avec un assistant connecté est immense et perd l'utilisateur dans son identification du produit. Lorsque nous utilisons une enceinte connectée pour mettre de la musique, nous ne l'envisageons pas de la même manière que lorsque nous lui demandons de nous faire une blague ou de nous dire qu'elle nous aime. La fonction musicale et toutes les fonctions « basiques » ramènent l'enceinte à son identité d'objet du quotidien alors que les fonctions conversationnelles ou plus complexes comme la gestion de l'agenda ou d'achat créent une proximité avec l'utilisateur parce qu'elles touchent à son humour ou à ses affaires personnelles. Dès lors, la relation avec l'enceinte connectée peut varier et nous ne lui attribuons pas toujours la même place dans notre quotidien. Les marques, en créant des dialogues propres à une personnalité réfléchie en amont donnent un corps et une identité à leurs enceintes connectées et jouent avec les différentes perceptions que peuvent avoir les utilisateurs du produit. À travers leurs publicités, les marques communiquent sur la place à attribuer à ces objets : pourquoi les utiliser, où les placer, comment les considérer. La place de cette technologie dans les foyers devient une question de société dans la mesure où celle-ci modifie les habitudes de consommation et de communication des citoyens. Dès lors, la sphère publique s'empare du sujet et des discours de sensibilisation à l'utilisation des enceintes connectées émergent. Pourtant, l'interface vocale n'est pas apparue récemment. Nous la connaissons dans nos téléphones portables, d'abord par nos messageries vocales puis par les assistants vocaux. Ainsi, dans quelle mesure les enceintes connectées ont-elles fait émerger un nouveau débat de société sur la place de la technologie ? Pourquoi, alors que les assistants vocaux sont connus du grand public depuis 2011, les enceintes connectées créent-elles une méfiance et accentuent-elles la peur de la surveillance de masse ? Nous analyserons dans cette seconde partie les discours liés à la place de l'enceinte connectée au sein du foyer, nous comparerons les messages des publicités des marques aux discours d'éducation préventive de la CNIL<sup>58</sup>. L'objectif sera ainsi de mettre

---

<sup>58</sup> Commission Nationale de l'Informatique et des Libertés, autorité administrative indépendante française.

en perspective deux visions de la technologie qui s'affrontent dans l'arène public : d'un côté des marques qui donnent une image bienveillante de l'assistantat numérique et de l'autre, l'État français qui sensibilise aux dangers de l'usage des données personnelles et à la surveillance de masse.

## A. L'enceinte connectée vue par les marques

### 1. Le *smart speaker* : une gouvernante 2.0

#### a. La professionnelle de l'organisation

En étudiant les publicités de Google, Amazon et Apple, nous remarquons que les enceintes connectées sont toujours, et ce sans exception, placées dans une pièce de vie commune de la maison : très souvent dans la cuisine, le salon ou la salle à manger. Selon l'étude de Capgemini, on retrouve cette pratique chez les usagers. En effet 69% des utilisateurs d'assistants vocaux les utilisent dans le salon et 61% dans la cuisine, ce qui correspond également à leur utilisation : les enceintes sont majoritairement utilisées en cuisinant, en *faisant autre chose* ou en regardant la télé selon une étude de PwC<sup>59</sup>.

L'enceinte connectée nous rappelle l'ordinateur des années 2000 : l'ordinateur fixe placé au cœur du salon, commun à tous et que chacun utilisait à tour de rôle. Le fonctionnement par l'oralité fait en effet de l'enceinte connectée un objet à partager et non à individualiser. Comme l'ordinateur fixe familial, l'enceinte peut regrouper plusieurs comptes en retenant différentes voix. Avec un Google Home, il est possible (aux États-Unis pour le moment) de distinguer six voix différentes et permettre ainsi d'enregistrer jusqu'à six différents comptes Google. L'enceinte peut ainsi répondre personnellement à un usager, lui donner le détail de sa journée et lire les mails d'un autre usager en reconnaissant le changement de voix.



À gauche, les sessions sur Windows XP et à droite les sessions d'un Google Home

L'usage des différentes sessions sur un même ordinateur s'était effacé avec le déploiement des smartphones et des ordinateurs portables dont l'usage est individuel plus que collectif. L'enceinte connectée fait renaître le partage d'un objet électronique et les marques

---

<sup>59</sup> Plus d'informations en annexe 3

insistent fortement sur ce point dans leurs communications. En étudiant les publicités des fabricants, nous observons que l'enceinte est présentée comme un nouveau membre de la famille. Les films représentent des moments de la vie familiale. L'enceinte est présentée comme la nouvelle gouvernante de la maison : celle qui n'oublie jamais les rendez-vous, a réponse à toutes les questions des enfants et peut même leur raconter des histoires pour les endormir. Nous remarquons par ailleurs que dans 100% des vidéos étudiées, qu'elles soient à destination d'un public anglophone ou francophone, les *smart speaker* ont tous des voix de femme.

L'aide aux parents devient une promesse de vente. Les mises en scène d'assistance aux parents débordés dans les publicités sont récurrentes. Que ce soit chez Amazon ou Google, l'assistant vocal semble devenir nécessaire à la vie parentale qui demande souvent d'être en situation de *multitasking*. L'enceinte connectée apparaît alors comme une aide-ménagère qui vient libérer les parents.

Dans une des vidéos<sup>60</sup> de la série « Google Home – profitez du moment », un père joue avec ses quatre enfants. N'ayant pas les mains libres, il se débarrasse des enfants avec sa Google Home en déclenchant sa télé à distance et en lançant un dessin animé.

Une publicité d'Amazon<sup>61</sup> va plus loin en disant clairement qu'Alexa « becoming part of the family » et ajoute le « témoignage » d'une utilisatrice « The Echo is a tool that we use to keep our household functioning<sup>62</sup> ». L'enceinte peut en effet contenir les emplois du temps de chacun, participer aux devoirs des enfants, gérer la domotique et aider lorsque les mains sont occupées à autre chose. Dans une autre publicité d'Amazon nommée « Dad's day »<sup>63</sup>, une jeune mère laisse à son compagnon leur bébé. Celui-ci semble devoir s'en occuper pour la journée et ne semble pas tout à fait à l'aise. Heureusement, avant de quitter la maison, la mère a programmé des *reminders* : l'anneau de dentition dans le frigo, le biberon à telle heure, le rendez-vous médical... Alexa est ainsi programmée pour rappeler au père ce qu'il doit faire heure par heure et l'aider dans sa mission. À la fin de la vidéo, le père exténué se repose lorsqu'Alexa lui dit : « I'm reminding you : Laure loves you and you are doing a great job. ». Ces deux exemples mettent en scène des parents débordés et soulagés par la présence de leur enceinte. L'enceinte connectée joue clairement le rôle de la gouvernante de maison tout en cherchant à décomplexer les parents débordés.

---

<sup>60</sup> Lien vers la vidéo : <https://www.youtube.com/watch?v=luhNAdNOpg4>

<sup>61</sup> Lien vers la vidéo : <https://www.youtube.com/watch?v=FQn6aFQwBQU&t=4s>

<sup>62</sup> « Echo est un outil que nous utilisons pour garder la maison en ordre »

<sup>63</sup> Lien vers la vidéo : <https://www.youtube.com/watch?v=qux0zs2jFgc>

Par ses mises en scène, les fabricants d'enceintes connectées nous livrent leur vision de la place des *smart speaker* dans nos vies. Sans qu'elle ne devienne un objet intime comme un smartphone, l'enceinte connectée est là pour gérer le quotidien de la communauté plus que la vie d'une seule personne. Elle semble devoir gérer une autre dimension de nos quotidiens, nos vies familiales et sociales – et ce hors des écrans.

#### b. L'oreille attentive pour les moments difficiles

Au-delà de son rôle organisationnel, l'enceinte connectée est représentée comme une gouvernante moderne aussi au travers de la voix qui lui donne une personnalité rassurante voire la représente comme une figure maternelle.

L'imaginaire commun met souvent en scène les personnes isolées avec un fond sonore : avec une radio ou une télévision toujours allumée, comme s'ils voulaient oublier la solitude et ne surtout pas entendre le silence angoissant. La voix, plus que la vue ou le toucher est un élément important dans nos vies sociales. Il permet de communiquer et le son d'une respiration ou un fond sonore d'une voix sourde nous permet de mieux apprécier une présence qu'une photo ou vidéo sans son.

L'enceinte connectée, même si elle peut tenter de parler naturellement de la pluie et du beau temps, apparaît surtout comme un soutien moral grâce à ses différentes fonctions et la présence de la voix. De plus, le fait que l'objet ait un « prénom » renforce l'ambiguïté de l'utilisation. Pour illustrer les propos précédents, nous nous baserons sur deux films publicitaires, l'un d'Apple et l'autre d'Amazon.

Dans le film d'Apple<sup>64</sup>, une jeune femme fatiguée prend le métro bondé, sort du travail sous la pluie et rejoint son appartement, déprimée. Son environnement est morose et gris, et parmi la foule, la jeune femme nous fait comprendre qu'elle se sent terriblement seule. Arrivée chez elle, elle demande alors à son Home Pod, avec une voix désespérée et triste « Hey Siri, play me something I like ». L'enceinte répond simplement un « Ok » avec une voix enjouée et lance une musique. Progressivement, la jeune femme se met à danser et tout son environnement se met en mouvement avec elle. À l'image de la marque qui se veut être le fournisseur officiel des « créatifs », le film est riche en créativité, la jeune femme danse et le décor se dédouble pour devenir sur-réaliste et déjanté. À la fin, le protagoniste semble se réveiller d'un rêve, allongée sur son canapé, détendue et souriante. La dernière prise du film :

---

<sup>64</sup> Lien vers la vidéo : <https://www.youtube.com/watch?v=305ryPvU6A8>

l'enceinte en gros plan qui apparaît comme la solution magique. Seules deux phrases sont prononcées dans cette publicité, lorsque le personnage demande à l'enceinte de mettre de la musique et que celle-ci lui répond « Ok ». On peut interpréter ce minimalisme comme une représentation de l'amitié : l'enceinte apparaît comme une personne qui connaît son utilisateur. Le film laisse imaginer que l'objet sent que la jeune femme ne va pas bien mais que ce ne sont pas les mots qui la réconforteront mais juste une présence bienveillante. L'actrice dit simplement « Hey Siri, play something I like » avec une moue triste – et l'enceinte lance une musique, faisant croire que celle-ci avait deviné ce qui ferait plaisir à sa protégée. Il s'agit ici probablement de la musique la plus écoutée sur son compte Apple Music. L'enceinte, bien qu'elle soit ici très discrète, apparaît comme un soutien moral en faisant simplement acte de présence. Cette publicité la montre très peu et pourtant en regardant la vidéo, on arrive à donner à l'enceinte une personnalité : une femme douce et attentive.

On retrouve le même cas de figure chez Amazon. Pour la sortie d'Amazon Echo en France, la marque a sorti un film plaçant encore une fois l'enceinte comme une « nounou » 2.0, à qui l'on confierait ses petits soucis. La vidéo met en scène une adolescente qui n'arrive pas à avouer quelque chose à ses parents. On comprend rapidement que celle-ci est homosexuelle et souhaite leur dire. Ses parents de leur côté viennent d'acheter une enceinte connectée et apprennent pas à pas à s'en servir. Tout au long du scénario, on voit la famille demander différentes choses à Alexa, dans les moments insignifiants du quotidien ou dans les instants stressants, comme celui qui précède un rendez-vous romantique. Une relation particulière se crée entre chaque personnage et le *smart speaker*, tout en laissant toujours l'objet en arrière-plan. À la fin, la jeune fille trouve enfin un moyen de faire son *coming-out*. Celle-ci arrive dans la cuisine où sont ses parents et demande à Alexa « Alexa, recommence ma musique » et l'enceinte lance la célèbre chanson de Jacques Dutronc : « J'aime les filles ». Le slogan final : « À vos côtés pour tous les petits moments de la vie, et les grands aussi ». L'enceinte est encore montrée comme une amie fidèle et encourageante. Comme chez Apple, le produit est peu montré mais la relation entre l'objet et les utilisateurs se fait beaucoup ressentir.

## 2. La simulation d'une relation de confiance entre utilisateur et fabricant

Selon l'étude de Capgemini, 65% des freins à l'adoption aux assistants vocaux sont liés aux problèmes de sentiments d'insécurité et 59% sont liés à l'aspect intrusif de l'objet et le fait qu'il enregistre beaucoup d'informations personnelles<sup>65</sup>. Face aux discours alarmants qui émergent au sein de l'arène public, les marques misent sur les aspects rassurants des enceintes connectées pour transmettre une vision bienveillante de la technologie. En valorisant certaines fonctions ou en simulant une écoute attentive des retours client, les marques laissent paraître une relation transparente entre utilisateur et enceinte connectée. Nous verrons dans cette seconde partie quels procédés communicationnels utilisent les marques pour s'armer contre les discours alarmants et introduire une relation de confiance.

### a. La conception d'un produit rassurant

Les marques développent un discours rassurant en travaillant particulièrement sur l'image de la technologie au sein des foyers pour tenter de la dé-diaboliser. En lui donnant une personnalité attentive et organisée, l'objet se veut comme l'ancienne gouvernante de maison, une aide indispensable à la vie familiale.

Au-delà d'un travail d'image, les fabricants ont conçu et valorisé les enceintes d'une certaine manière de façon à aller à l'encontre des discours anti-assistants vocaux. Nous verrons dans cette partie comment le design et la valorisation de certaines fonctions servent le discours des fabricants et rassurent le consommateur.

Pour illustrer ces propos, nous prendrons l'exemple d'une publicité Google sortie en 2016<sup>66</sup>. Le film présente uniquement le produit : on y voit seulement un Google Home sur un fond blanc. Google considère ici que son auditoire est familier des fonctions d'une enceinte connectée puisque le film ne fait aucune démonstration et n'évoque pas les fonctionnalités du Google Assistant. Au contraire, la première fonction montrée est l'écran tactile « Clean touch pad surface for total control ». Ce « touch pad » répond aux demandes suivantes :

- Lancer, mettre en pause ou arrêter l'alarme ou le minuteur
- Augmenter le volume
- Réduire le volume

---

<sup>65</sup> Plus d'informations en annexe 3

<sup>66</sup> Google, *Introducing Google Home*, 04.10.2016 - lien vers la vidéo : <https://www.youtube.com/watch?v=r0iLfAV0pIg>

- Lancer votre demande
- Désactiver le micro
- Rétablir la configuration de l'appareil<sup>67</sup>

L'interface tactile a donc peu d'intérêt au vu des nombreuses autres fonctions permises par une enceinte connectée. Le fabricant cherche en fait à rassurer la cible quant au contrôle que celle-ci aura sur son objet en rappelant explicitement « Clean touch pad for total control ». Pourtant l'interface tactile n'est pas vraiment une interface graphique puisqu'il s'agit de simplement de led colorées qui indiquent un *feedback* à l'utilisateur (quand celui-ci monte le son, les lumières s'allument et suivent son geste pour montrer que la machine est bien en train de s'exécuter). Cette surface tactile ne permet donc en aucun cas un *contrôle total* sur la machine d'autant plus que ses fonctions ne sont pas explicites. Il n'y a aucun signifiant pour savoir comment baisser le son par exemple. Cette fonction doit s'apprendre par la notice d'utilisation. Cet écran est donc loin d'être la fonction la plus utile et la plus utilisée parce qu'elle nécessite un certain apprentissage et qu'elle n'a finalement que peu d'intérêt. Mais puisque la perte de l'interface graphique est perturbante pour l'utilisateur, Google cherche ici à combler un manque et à rassurer en donnant une impression de contrôle sur la machine par le toucher et un feedback visuel, deux éléments bien connus par le consommateur.



*Extrait de la vidéo Introducing Google Home*

---

<sup>67</sup> Capture d'écran de la notice en annexe 2

La seconde fonctionnalité mise en avant (à ce stade du film, on ne sait toujours pas à quoi sert concrètement l'enceinte) est la reconnaissance vocale. Les deux petits traits noirs aux extrémités de l'enceinte sont les micros qui enregistrent la voix de l'utilisateur. Le film fait un rapide passage sur ceux-ci avec la légende : « Far-field voice recognition for hands-free use ». Alors que l'on passe cinq secondes sur la première fonction, le « touch pad », on ne passe que 3 secondes sur la reconnaissance vocale, pourtant première fonctionnalité du produit. L'usage de l'enceinte n'est donc pas le sujet de cette publicité censée présenter le produit mais un moyen de rassurer le consommateur sur les peurs identifiées : la perte de l'absence d'interface graphique et le manque de respect de la vie privée. Dès lors, la troisième fonction valorisée est le bouton *mute* : « Mic mute button for privacy ». On fait un zoom sur ce bouton pendant 4 secondes pour rappeler que si l'enceinte écoute et comprend, elle peut aussi être désactivée et activée, et ce par le toucher, grâce à un bouton à presser. Ce dernier est composé de trois signifiants : sa forme ronde soulignée par un contour, son relief embossé qui suggère une action et le pictogramme « micro barré » qui indique la fonction du bouton. Trois signifiants pour le bouton *mute* et aucun pour le reste de l'appareil : la fonction d'arrêt du micro est en effet valorisée et facile à comprendre pour rassurer le consommateur.



Amazon et Alibaba placent le bouton *mute* sur la surface de l'enceinte pour encore plus de visibilité. Xiaomi a intégré un bouton « On/Off » pour éteindre l'enceinte<sup>68</sup>. Ces constructeurs ont ainsi misé sur les angoisses du manque de contrôle liées à l'interface vocale.

---

<sup>68</sup> Photos des différentes enceintes en annexe 4

L'HomePod d'Apple ne comporte pas de bouton *mute*, il est donc impossible de couper l'enceinte à moins de la débrancher, un aspect qui peut se révéler anxiogène.

La majorité des fabricants ont identifié deux peurs : la perte du contrôle de la machine par manque de contact visuel et l'intrusion dans la vie privée. En réponse, les design des enceintes ressemblent à ce que le consommateur connaît : une surface tactile ou des boutons. Malgré le peu d'utilité des boutons centraux, les marques n'osent pas la surface plate, sans bouton ni interface. Même Apple, qui se différencie des autres marques par l'absence de bouton *mute*, a intégré une surface tactile à l'HomePod. L'absence de bouton *mute* chez Apple peut d'ailleurs être interprétée comme un message de confiance : pourquoi intégrer un bouton *mute* si l'enceinte ne vous écoute jamais, à part quand vous avez besoin d'elle ? Ce n'est pas le parti pris des autres fabricants qui valorisent au contraire le bouton *mute*, que ce soit par le design ou à travers la communication – pour rappeler à l'utilisateur sa liberté, son contrôle sur sa connexion.

b. Les prises en compte des retours client : stratégie de communication ou réelle prise de conscience ?

Nous avons vu dans les parties précédentes que les enceintes connectées avaient participé au débat sur l'intrusion des entreprises Tech dans nos vies privées. Lors de sa dernière conférence annuelle, Google avait annoncé souhaiter guider ses utilisateurs pour une meilleure utilisation de leur smartphone en les aidant à limiter leur usage. Comme nous l'avons vu dans la première partie, il est aujourd'hui prouvé que la surconsommation de smartphone rend son possesseur angoissé et augmente la probabilité de dépression. L'angoisse d'être séparé de son téléphone a aujourd'hui un nom, la nomophobie, « **no mobile-phone phobia** ». Cette phobie n'est pas inscrite dans le « manuel diagnostique et statistique des troubles mentaux<sup>69</sup> » et n'est donc pas reconnu par la communauté scientifique internationale comme étant un trouble mental, contrairement à l'addiction aux jeux d'argent. Pourtant, en Chine cette addiction est considérée comme un trouble clinique depuis 2008 – et comme la première menace de santé pour les adolescents. En réponse à ce diagnostic, le pays asiatique a ouvert plusieurs centres de désintoxication spécialisés. Dans un documentaire du

---

<sup>69</sup> Le *Manuel diagnostique et statistique des troubles mentaux* (également désigné par le sigle *DSM*, abréviation de l'anglais : *Diagnostic and Statistical Manual of Mental Disorders*) est un ouvrage de référence publié par l'Association américaine de psychiatrie (*American Psychiatric Association* ou *APA*) décrivant et classifiant les troubles mentaux.

New-York Times<sup>70</sup> réalisé en 2014, les journalistes partent en reportage dans un de ces centres. Les bâtiments, les chambres minuscules, les barreaux aux portes, tout ressemble à une prison. Les pensionnaires semblent avoir entre 10 et 16 ans et sont tous des anciens « addicts à Internet ». Placés par leurs parents, ils passent trois ou quatre mois en désintoxication, suivant un rythme de vie militaire et une thérapie. Le monde occidental tolère davantage les nouvelles technologies, et même si elles reconnaissent leurs torts, aucun plan de désintoxication n'est envisagé. La distance avec les nouvelles technologies est perçue comme une affaire individuelle. Chacun doit apprendre à gérer ses comportements, et en face, les entreprises doivent être transparentes. En Chine, la dépendance à Internet est considérée comme une véritable maladie à vaincre par la discipline et le contact humain. Si les deux visions sont différentes, il est déjà acquis que l'addiction à Internet est un véritable problème de santé publique.

Les GAFA réagissent progressivement face à cette alerte générale. Google a mis en place un *dashboard* sur la dernière version d'Android qui permet plusieurs fonctionnalités : la suspension des notifications à certaines heures, la possibilité de limiter son temps sur certaines applications, et un tableau de bord qui résume notre consommation de la journée (1 heure sur l'application Facebook et 34 minutes sur Instagram, 23 messages texte envoyés, etc.). Ce *dashboard*, tableau de bord en français, est vendu comme un assistant qui a pour mission de limiter notre addiction. Les GAFA nous guident une fois de plus dans notre consommation numérique, nous encourageant à limiter notre temps sur les applications, à cesser de réagir à la présence de chaque notification. Les GAFA considèrent-ils les utilisateurs comme des idiots irresponsables ? L'apparition du *dashboard* n'est pas une manière de mater des consommateurs incapables de se contrôler mais est une anticipation à l'évolution des usages. Les GAFA réalisent progressivement les limites de leurs produits mais ne veulent pas perdre le contrôle sur leurs milliards d'utilisateurs. Le *dashbaord* est un exemple de prise de pouvoir de Google en anticipation à un changement d'habitude de consommation. En permettant aux utilisateurs de régler ses consommations, d'éteindre les notifications à une certaine heure, etc. Google collecte encore plus de données sur leurs habitudes.

Les entreprises de la Tech évoluent en permanence avec leurs usagers en particulier parce que leurs produits peuvent être mis à jour par le biais de leurs serveurs. Par exemple, en

---

<sup>70</sup> Lien vers le documentaire du New-York Times, réalisé par Shosh Shlam et Hilla Medalia, sorti le 20 janvier 2014 : <https://www.nytimes.com/video/opinion/100000002657962/chinas-web-junkies.html?playlistId=1194811622299>

2016, un premier article sort sur le sujet « Amazon Echo Is Magical. It's Also Turning My Kid Into an Asshole. ». L'auteur est un parent, ancien employé de Google, qui tient un blog sur le numérique<sup>71</sup>. Le père s'alarme de voir sa fille adorer Alexa et lui donner des ordres. Il a peur qu'un enfant ne comprenne pas pourquoi il peut parler d'une telle manière à Alexa et non à une personne. D'autant plus qu'il souligne que ces objets sont très appréciés des enfants qui ne savent pas encore lire et écrire et pour qui la communication passe principalement par la voix. Il conclut ainsi « So Amazon, you clearly have a hit on your hands. Can I request one thing? A kid-mode where the Echo only responds to "Alexa, please..." as opposed to just "Alexa." ». La journaliste Hélène Chevallier s'est intéressée au phénomène et met en parallèle les plaintes des parents avec les travaux du psychiatre Serge Tisseron qui se veut finalement rassurant en rappelant que les enfants testent avec les objets des comportements qu'ils ne peuvent avoir avec les adultes ou les autres enfants. C'est pour cela qu'il ne serait pas rare de voir un enfant être très autoritaire avec ses poupées ou avec une enceinte connectée, tout en respectant les règles de politesse avec les adultes. Pourtant, en mai 2017, lors de sa conférence I/O, Google introduisait le « Pretty please ». Alors que le VP Engineering de Google, Scott Huffman, liste les nouveautés de Google Home, il confie à l'assemblée : « A concern I heard from many parents, including the people on the team who have children is kids are learning to be bossy and demanding when they can just say 'Hey Google' to ask for anything they need.<sup>72</sup> » Il explique ensuite que certains parents de ses équipes ont pu tester l'option et lance une vidéo mettant en scène parents et enfants en interaction avec Google Home. Avec cette nouvelle option, les enfants peuvent formuler leurs demandes autrement, en disant « please » et Google Assistant répond en conséquence :

« *Ok Google, please tell me a story.* » → « *Thanks for saying please.* » / « *Thanks for asking so nicely.* »

« *Please, help me with my homework* » → « *You're very polite.* »

« *Ok Google, play Freeze Dance please* » → « *What a nice way to ask me.* »

---

<sup>71</sup> Walk Hunter, 06.04.2016 : <https://hunterwalk.com/2016/04/06/amazon-echo-is-magical-its-also-turning-my-kid-into-an-asshole/>

<sup>72</sup> Lien vers la vidéo de la conférence I/O 2018 : <https://www.youtube.com/watch?v=ogfYd705cRs&t=3861s> (à partir de 26''35) – traduction « J'ai eu des retours d'inquiétudes de la part d'un grand nombre de parents, y compris les membres de l'équipe qui ont des enfants. Les enfants deviennent autoritaires et exigent beaucoup quand ils peuvent simplement dire «Hey Google» pour demander n'importe quoi. »

L'option est activable sur l'application Google Assistant, l'enfant doit ensuite enregistrer plusieurs fois sa voix pour que l'enceinte la reconnaisse et complexifie son *wake word* de « Ok Google » à « Ok Google + élément + please » ou « Ok Google, please ». L'enceinte connectée n'est plus vue comme un objet réservé aux adultes mais comme adaptée aux enfants. En les encourageant à dire « please », l'enceinte apparaît comme un support, voire un complément à l'éducation apportée par les parents. « Pretty please » les confortent dans l'idée qu'une Google Home peut les aider à éduquer des enfants polis et ainsi simplifier leur quotidien.

## **B. L'émergence d'un débat sur la protection des données personnelles**

Le 21 mars 1974, le quotidien *Le Monde* révèle que le gouvernement souhaite lancer le projet SAFARI<sup>73</sup> ayant pour objectif d'identifier chaque citoyen par un numéro et d'interconnecter tous les fichiers de l'administration. Cette révélation créa une vive émotion dans l'opinion publique. Cela a conduit le gouvernement à créer une commission afin qu'elle propose des mesures garantissant que le développement de l'informatique se réalise dans le respect de la vie privée, des libertés individuelles et publiques. Cette commission proposa de créer une autorité indépendante pour réguler le développement d'Internet. Le 6 janvier 1978, la loi relative à l'informatique, aux fichiers et aux libertés institua ainsi la Commission Nationale de l'Informatique et des Libertés (CNIL). La CNIL se compose de 18 membres et d'une équipe d'agents contractuels de l'Etat. Selon le site officiel de la CNIL, l'autorité administrative a quatre missions principales :

- Informer, protéger les droits
- Accompagner la conformité / conseiller
- Anticiper et innover
- Contrôler et sanctionner <sup>74</sup>

En Europe, l'ensemble des « CNIL européennes » se nomme le G29. Le groupe se réunit une fois tous les deux mois environ à Bruxelles pour « contribuer à l'élaboration des normes européennes en adoptant des recommandations, rendre des avis sur le niveau de protection dans les pays hors UE et conseiller la Commission européenne sur tout projet ayant une incidence la protection des données et des libertés des personnes.<sup>75</sup> ». La directive européenne du 24 octobre 1995 constitue le socle commun à tous les pays de l'Union européenne en matière de protection des données personnelles. Le texte a pour objectif de lisser l'ensemble des législations nationales sur la protection des données afin de permettre une cohérence entre les différents pays et simplifier les procédures en cas de litige. En 2012, Bruxelles a proposé de mettre à jour la loi de 1995 pour tenir compte des évolutions des technologies, c'est ainsi que le projet du RGPD<sup>76</sup> vu le jour. Nous verrons dans un premier

---

<sup>73</sup> Système Automatisé pour les Fichiers Administratifs et le Répertoire des Individus.

<sup>74</sup> Selon le site : <https://www.cnil.fr/fr/missions-et-fonctionnement-de-la-cnil>

<sup>75</sup> Selon le site : <https://www.cnil.fr/fr/le-g29-groupe-des-cnil-europeennes>

<sup>76</sup> Règlement Général sur la Protection des Données

temps quel est le positionnement global de l'Europe sur la protection des données. Puis, il conviendra d'analyser le discours de la CNIL au sujet des enceintes connectées.

## 1. En Europe : conséquences de la mise en place du RGPD

Le Règlement Général sur la Protection des Données (RGPD) a pour objectif de réguler, à échelle européenne, le traitement et la circulation des données à caractère personnel. Le RGPD a été en discussion de 2012 à 2016, année durant laquelle le texte définitif a été validé. Le texte s'est ensuite appliqué à partir du 25 mai 2018.

### a. Qu'est-ce qu'une donnée personnelle ?

Selon le glossaire du site de la CNIL, une donnée personnelle est « toute information identifiant directement ou indirectement une personne physique (ex. nom, no d'immatriculation, numéro de téléphone, photographie, date de naissance, commune de résidence, empreinte digitale...)»<sup>77</sup> ». Certaines des données peuvent être sensibles car pouvant permettre de discriminer un usager. Il peut s'agir d'opinion politique, religieuse, un engagement syndical, une appartenance ethnique, une orientation sexuelle, une situation médicale ou des idées philosophiques. Ces données sont régies différemment. En effet, il est impossible pour une entreprise de collecter une de ces données sans l'autorisation écrite, claire et explicite de la personne concernée.

### b. Les principes du RGPD

Le RGPD s'applique à toute entreprise souhaitant utiliser des données européennes, peu importe l'origine de l'entreprise, son lieu de stockage ou son domaine d'activité. Son objectif est de protéger les données du citoyen et d'empêcher les entreprises de les utiliser selon leurs propres règles. Par exemple, le RGPD a mis en place le principe de co-responsabilité. Pour être en règle, une entreprise doit s'assurer de la conformité de ses fournisseurs vis-à-vis de la réglementation. Cela empêche un acteur de se dédouaner d'une mauvaise utilisation de données en passant par un fournisseur ou partenaire externe. Aussi, le RGPD doit être implanté dès la phase de conception d'un projet, c'est ce que l'on appelle le *privacy by design*. Ainsi, une entreprise doit prendre en compte l'ensemble du règlement en sécurisant dès le départ les données personnelles et en étant totalement transparent avec

---

<sup>77</sup> Selon le site : <https://www.cnil.fr/fr/glossaire>

l'utilisateur. L'entreprise doit également garantir par défaut le plus haut niveau possible de protection des données, et ce de manière continue.

## 2. Enfants, invités, vie privée : analyse des conseils de la CNIL pour une bonne gestion de son enceinte connectée

Lorsque nous utilisons Google ou regardons Youtube, la requête est envoyée dans leur serveur. Nous sommes également suivis lorsque nous sommes sur Facebook ou sur un site qui nécessite un compte du réseau social. La majorité de nos habitudes numériques sont aujourd'hui liées aux GAFA. Le RGPD régule depuis mai 2018 cette collecte des données en imposant une demande d'autorisation et une explication claire de l'utilisation des données personnelles à l'internaute. Ces nouvelles règles sont aujourd'hui faciles à intégrer lorsqu'il y a un écran mais qu'en est-il de la sécurité avec les interfaces vocales ? Comment la CNIL réagit-elle face à l'intégration d'une nouvelle interface qui complique la demande d'autorisation et renforce la surveillance de masse ?

Il s'agit en effet d'un nouvel enjeu pour l'instance dans la mesure où cette nouvelle interface instaure de nouveaux usages. En effet, l'enceinte vocale, parce qu'elle enregistre n'importe quelle voix ayant prononcé le *wake word* demandé, touche un cercle beaucoup plus large d'utilisateurs. L'étude de Capgemini sur les usages des enceintes connectées nous montre que les utilisateurs placent leur *smart speaker* majoritairement dans leur cuisine ou leur salon, des pièces de vie commune. L'enceinte connectée n'est pas donc perçue comme un objet personnel mais plutôt un bien commun. Son utilisation, bien qu'elle ne fonctionne qu'en *one-to-one*, est une expérience partagée aux oreilles de tous ceux qui sont dans les alentours. Il devient ainsi difficile de priver quelqu'un d'utiliser l'enceinte connectée, chacun participant à l'expérience de l'autre par le lien de l'oralité.

La CNIL a publié en décembre 2017 une page web dans laquelle elle explique le fonctionnement d'une enceinte connectée et développe son point de vue sur les enjeux de confidentialité que pose la machine. La CNIL a d'abord défini quatre grandes étapes de fonctionnement d'une enceinte connectée<sup>78</sup> : réveiller l'enceinte / énoncer sa demande / retranscription de la parole en texte puis interprétation et réponse / remise en veille de

---

<sup>78</sup> D'après la page web : <https://www.cnil.fr/fr/enceintes-intelligentes-des-assistants-vocaux-connectes-votre-vie-privée>

l'enceinte. Cette page web est une notice d'utilisation qui prétend à apprendre au lecteur comment fonctionne une enceinte connectée tout en le sensibilisant sur ce qu'implique l'objet en terme d'intrusion dans sa vie privée. La CNIL joue un rôle de médiateur entre la technologie et l'utilisateur en simplifiant son discours et en adaptant un ton pédagogique. La page est aérée, bien structurée et tous les propos sont illustrés. Les dessins colorés et l'écriture manuscrite nous ramènent dans l'univers scolaire.

**Première étape :** « L'utilisateur « réveille » l'enceinte à l'aide d'une expression clé (« Hey Snips » / « Ok Google » / « Hey Alexa »). »

*Ci-dessous l'image d'illustration :*



**Deuxième étape :** « L'utilisateur énonce sa requête. Certaines de ces enceintes enregistrent localement les requêtes de l'utilisateur de manière à lui laisser la maîtrise de ses données (ex. une enceinte connectée avec l'assistant vocal de Snips). D'autres en revanche, envoient ces requêtes dans le cloud, autrement dit sur les serveurs de traitement de la société (ex. Amazon Echo, Google Home, ...). » Nous notons ici, comme dans la première étape, que la CNIL cite Snips – un créateur d'assistants vocaux intégrés aux objets. L'entreprise, française, vend aux fabricants de toutes sortes d'objets des petits ordinateurs à intégrer pour faire les parler. Par exemple, un fabricant de machine à café peut acheter à Snips un assistant vocal pour l'intégrer sur une machine. L'assistant vocal ne dépend d'aucun serveur, les données sont ainsi la propriété des utilisateurs et ne sont en aucun cas partagées. Snips est une entreprise française dirigée Hand Hindi, défenseur de la protection des données personnelles. Dans une interview pour Capitale le fondateur déclarait :

*« La confidentialité devient [...] un enjeu capital car le recueil de ces données, et ce qu'on en fait, touche à la vie privée des utilisateurs. Il faut avoir une position claire. Facebook a perdu 90 milliards de dollars à cause de son flou sur la question. Aujourd'hui, le principal frein à l'adoption des objets connectés, c'est la peur du "Big Brother" californien. »*

Yann Lechelle, COO de Snips, déclarait dans une interview pour Libération, que la voix est un marqueur biométrique, comme l'empreinte digitale, et n'a donc pas à être stockée dans un "cloud". Libération évoque le soutien de la CNIL dans cette initiative. Gwendal Le Grand, Directeur des technologies et de l'innovation à la Cnil, se réjouit de voir que certaines entreprises se positionnent sur la protection de la vie privée, cela montre effectivement que « le règlement européen sur la protection des données peut favoriser l'innovation »<sup>79</sup>. Lors de notre entretien, Yann Lechelle déclarait en effet que « L'ADN de Snips c'est le *Privacy by design*.<sup>80</sup> ». Snips se positionne comme étant une solution pour faire reculer le pouvoir des GAFAs et empêcher l'Europe de devenir une « colonie du monde numérique<sup>81</sup> ». En effet, l'écosystème numérique mondial est aujourd'hui centralisé aux Etats-Unis. Sur 161 licornes – entreprises non cotées en bourse mais dont la valorisation dépasse un milliard de dollars – qui opèrent dans le numérique, 93 sont situées aux Etats-Unis, 38 en Asie et 15 en Europe<sup>82</sup>. Gilles Babinet, "Digital Champion" de la France auprès de la Commission européenne explique le monopole des Etats-Unis :

*« Le succès du numérique aux Etats-Unis est lié à quatre principaux facteurs d'émergence : un accès abondant au capital, un système universitaire de qualité et capable de collaborer avec le monde de l'innovation, des écosystèmes de taille critique mêlant start-ups, entreprises et universités polarisées sur le sujet du numérique, et enfin politiques publiques accommodantes (interdiction de taxes des Etats fédéraux sur l'e-commerce par exemple).<sup>83</sup> »*

Le retard numérique de l'Europe serait dû à un manque d'investissements dans ce secteur et de cohésion entre les pays de l'Union Européenne<sup>84</sup>. Les instances publiques

---

<sup>79</sup> Christophe Alix et Amaelle Guiton, « «Assistants vocaux» : des dispositifs très écoutants et peu regardants », 17 juin 2018, Libération.fr

<sup>80</sup> Entretien avec Yann Lechelle, COO de Snips, annexe 1

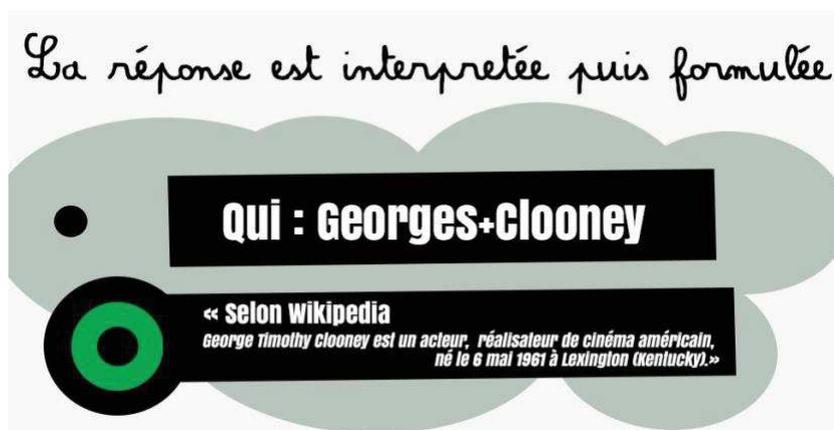
<sup>81</sup> Le terme « colonie du monde numérique » est utilisé dans un rapport d'informations du Sénat intitulé « L'Union Européenne : colonie du monde numérique ? » en 2010 : <http://www.senat.fr/rap/r12-443/r12-4432.html>

<sup>82</sup> Tobias KOLLMANN et Amal TALEB, « Comment faire de l'Europe un leader du numérique ? », confrontations.org, site de Confrontations Europe « think tank européen », 13.07.2016, consulté le 12.09.2018. <http://confrontations.org/interface/comment-faire-de-leurope-un-leader-du-numerique/>

<sup>74</sup> « Numérique : le retard européen », synthèse du site du Centre d'information sur l'Europe, 15.03.2018, consulté le 12.09.2018. <https://www.toutleurope.eu/actualite/numerique-le-retard-europeen.html>

regrettent aujourd'hui ce manque d'initiative européenne et française dans le domaine du numérique. Ce secteur est pourtant identifié comme porteur d'emplois et de croissance. La Vice-présidente du CNNum, le Conseil National du Numérique, Amal Taleb, considère que "l'éducation", "le financement" ainsi qu'un "cadre juridique clair" sont essentiels au développement du numérique européen. La mise en place du RGPD répond à l'ambition de l'Europe de vouloir rattraper son retard en s'imposant davantage et en développant des innovations plus adaptées à ses valeurs.

**Troisième étape :** « La parole prononcée est automatiquement transcrite en texte puis interprétée afin qu'une réponse adaptée soit fournie. »



*Image d'illustration du site de la CNIL pour la troisième étape*

L'illustration met en scène le fonctionnement technique d'un assistant vocal en écrivant « Qui : Georges + Clooney ». Cette image explicite le langage informatique utilisé par les logiciels et permet au lecteur de comprendre comment la machine analyse et ainsi comment lui parler.

**Quatrième étape :** « L'enceinte repasse en 'veille' »

À la suite de la description du fonctionnement de la machine, la CNIL introduit les enjeux liés à la vie privée des utilisateurs. Elle cible trois points clés qu'elle considère à expliciter : la confidentialité des échanges, la monétisation de l'intime et l'absence d'écran.

---

<sup>84</sup> *ibid.*

## La confidentialité des échanges

La CNIL rappelle que la technologie n'est pas parfaitement au point. L'enceinte peut lancer un enregistrement à l'insu d'un utilisateur à cause d'une mauvaise interprétation d'un mot qui ressemble trop au *wake-word*. L'enceinte connectée est un objet sensible qui peut se déclencher rapidement. La publicité de Burger King dans laquelle l'acteur demande « Ok Google, what is a Whopper ? » déclenchant ainsi le Google Home de tous les spectateurs en ayant une<sup>85</sup> est révélateur des limites de l'interface vocale. En réaction au dispositif de Burger King, Google a identifié la voix de l'acteur et pu stopper la publicité mais ne s'est ensuite jamais prononcé sur cette faille. Il y a donc une marge d'erreur à considérer et il ne faut pas oublier qu'on peut être enregistré à tout moment – à moins de ne couper le micro. La CNIL insiste également sur les enfants qu'il faut encadrer et ne pas laisser seul avec l'appareil. Bien que la loi européenne n'autorise pas les fabricants à concevoir un produit ou à mettre à disposition une application ayant pour cible un mineur de moins de treize ans, les parents peuvent les autoriser à les utiliser. Les enfants peuvent pourtant ne pas comprendre ce qu'ils font réellement et quelles traces ils laissent aux serveurs des fabricants et peuvent même acheter des articles. En réponse à ce problème d'achat provenant d'enfants, les *smart speakers* reconnaissent maintenant différentes voix et Google, par exemple, permet de créer un compte pour son enfant de moins de 13 ans qui aura par défaut un accès limité à l'assistant. Les enfants de moins de 13 ayant un compte Google synchronisé avec une Google Home ne peuvent pas lancer Youtube ou Youtube Music, Google Play Music (sauf s'ils ont l'option Google Play Music Family), faire des achats, utiliser des applications qui n'appartiennent pas à Google à moins que celles-ci aient la certification « Apps for families ». Amazon a même créé un Echo dédié aux enfants avec une version d'Alexa adaptée : reconnaissance plus précise des voix aigües des enfants et contrôle du temps d'utilisation et des contenus diffusés. La CNIL conseille alors de veiller à utiliser ces versions adaptées. Autre problème lié, les invités. Nous entendons par invité, toute personne qui n'habite pas le foyer propriétaire d'une enceinte connectée. La CNIL préconise de prévenir les invités de la présence d'une enceinte connectée et de les avertir que leur conversation pourra peut-être être enregistrée. S'il s'agit aujourd'hui d'une simple recommandation, nous pouvons d'ores-et-déjà imaginer qu'un litige opposant un propriétaire d'enceinte connectée et une personne enregistrée à son insu sera probablement à l'origine d'une révision de la loi européenne.

---

<sup>85</sup> Vidéo mettant en situation la publicité avec une Google Home : [https://www.youtube.com/watch?time\\_continue=19&v=n5lj63-nc5g](https://www.youtube.com/watch?time_continue=19&v=n5lj63-nc5g)

### **La monétisation de l'intime**

La CNIL prévient « [...] les appareils dotés d'un assistant à commande vocale se retrouvent au cœur de la vie du foyer. ». L'enceinte connectée va être envahissante car toujours à l'écoute, toujours présente. L'instance fait appel au bon sens de l'utilisateur et lui conseille de ne connecter à son enceinte que ce qui est vraiment utile. Le texte cherche à sensibiliser l'utilisateur en lui demandant d'être conscient des risques liés à la connexion d'objets trop sensibles à une enceinte connectée (les objets liés à la sécurité de la maison par exemple). Aussi, comme tout fichier numérique, ce qui est dit à une enceinte connectée sert à renforcer le profil publicitaire de son utilisateur. La CNIL le rappelle et considère qu'il faut être vigilant sur ce point. En d'autres termes, la CNIL nous avertit de ne pas dire n'importe quoi devant notre enceinte connectée. Au-delà de la collecte d'informations par les fabricants des assistants vocaux, les enregistrements peuvent être piratés et utilisés contre son utilisateur, tout comme une vidéo sur un portable ou un texto. L'instance rappelle que même si « la parole s'envole, les requêtes vocales sont enregistrées dans le cloud, de la même manière qu'elles le seraient s'il les tapait au clavier dans certains moteurs de recherche ». L'aspect éphémère et invisible de la requête vocale fait en effet oublier que tout ce que nous demandons après le *wake word* est enregistré.

### **L'absence d'écran**

La CNIL évoque également l'absence d'écran. Comme nous le disions précédemment, l'aspect éphémère et invisible de la requête vocale fait en effet oublier que tout ce que nous demandons après le *wake word* est enregistré. L'instance prévient donc :

*« L'individu échange vocalement avec un assistant connecté sans nécessairement s'appuyer sur un tableau de bord visuel. Sans écran, difficile d'avoir un aperçu des traces enregistrées, ni de juger de la pertinence des suggestions, d'en savoir plus ou d'avoir accès à des réponses provenant d'autres sources. »*

La CNIL recommande d'aller voir régulièrement son historique et de le supprimer ou de le personnaliser en fonction de nos besoins. L'instance rappelle que nous pouvons choisir notre moteur de recherche ou la source d'information utilisée, en somme, que nous sommes libres de l'utilisation que nous faisons de notre enceinte connectée. L'objectif ici est de rappeler à l'utilisateur que derrière cette interface vocale, il y a des traces écrites et la possibilité de les gérer. Encourager l'utilisateur à aller voir fréquemment l'historique de ses commandes vocales le pousse à prendre conscience que ses données sont bien réelles et

envoyées aux serveurs d'une entreprise dans l'objectif de les exploiter. Voir ses commandes est aussi un moyen pour l'utilisateur de réaliser ce qu'il dit et de réguler son comportement en fonction de ce qu'il veut partager avec le fabricant.

Enfin, la CNIL rappelle qu'elle est en contact avec les différents fabricants afin de comprendre comment évoluent ces objets. Les équipes de l'instance feraient des tests et mèneraient des réflexions pour garantir aux citoyens d'être informés au mieux de l'utilisation de leurs données personnelles.

## CONCLUSION DE LA PARTIE II

L'objectif de cette seconde partie était de comprendre dans un premier temps dans quelle mesure les discours promotionnels des enceintes connectées étaient porteurs d'une idéologie sur la place de la technologie dans la société. Le second objectif était de comprendre pourquoi les enceintes connectées faisaient l'objet de discours de sensibilisation de la part des États alors qu'elles ne sont finalement qu'un support supplémentaire aux assistants vocaux, des logiciels connus depuis 2011. Les assistants vocaux comme Google Assistant ou Siri sont pourtant les sources du problème soulevé dans la société actuelle puisque ce sont eux qui collectent les données pour affiner le profil publicitaire de chaque utilisateur. Notre objectif fut de comprendre pourquoi l'enceinte connectée centralise une majorité des critiques et des préventions contre la surveillance de masse.

Nous avons vu dans la première partie comment les marques communiquent sur leur vision de la place de la technologie dans les foyers. En vendant un objet collectif, les fabricants bousculent les habitudes numériques actuelles tournées vers un usage individuel et personnel. Si l'utilisation d'une enceinte connectée ne peut se faire que vers un utilisateur, il est possible de synchroniser un *smart speaker* avec plusieurs comptes. Une enceinte Google Home peut reconnaître jusqu'à six voix et ainsi personnaliser les réponses de chaque utilisateur. Les communications des marques placent toujours leurs produits dans un endroit central et commun du foyer comme la cuisine ou le salon. L'enceinte connectée est donc vendue comme élément central de la gestion de la maison. Nous avons vu que les marques l'humanisaient dans un rôle de gouvernante. Elle régit toute la maison, de la domotique aux emplois du temps de chacun. Elle aide les parents débordés et fait aussi acte de présence dans les moments difficiles. À la fois présentée comme point central du foyer, elle a donc aussi une

place particulière dans la vie de chacun. L'étude de PwC observe d'ailleurs que 38% des usagers préfèrent l'utiliser quand ils sont au lit contre 37% pour le smartphone. Le *smart speaker* entre peu à peu dans l'intimité de ses usagers. Peut-on imaginer que les usages vont évoluer vers une utilisation beaucoup plus individuelle ? En attendant les marques travaillent à donner une vision bienveillante voire maternelle de la technologie pour affronter les peurs émergentes liées à la surveillance de masse. Nous observons ainsi que les fabricants mettent l'accent sur d'autres aspects des enceintes pour en faire la promotion. L'exemple de Google illustre cette volonté de rassurer sur la vision du produit. La marque valorise le « touch pad », la partie de l'enceinte tactile et le bouton permettant de couper le micro – deux fonctionnalités qui ne sont aucunement les premières fonctionnalités du produit. Les marques adaptent leurs discours. Elles rassurent l'utilisateur, lui confirment que celui-ci aura toujours le contrôle qu'il pense avoir perdu à cause de l'absence d'interface graphique. L'enceinte connectée d'Apple est limitée puisqu'on ne peut demander que très peu de choses à Siri. Apple a pour l'instant pris le parti de ne vendre qu'une « enceinte augmentée » que l'on peut commander de loin pour simplifier l'usage. La gestion de la vie personnelle n'est ainsi pas une fonction permise par l'Home Pod – pourtant équipé d'une version de Siri. On note ainsi la démarche prudente des fabricants qui développent leurs produits en parallèle d'une prise de conscience émergente de la sensibilité des données personnelles et d'une réflexion générale autour de la place à donner aux nouvelles technologies. Nous avons ainsi analysé l'évolution des enceintes connectées avec le cas du « pretty please » et du *multitasking* de Google. La firme développe un autre type de discours rassurant, celui de l'évolution des produits en fonction des retours clients et de l'évolution des usages. En permettant aux enfants de dire « s'il te plaît » aux enceintes connectées ou en améliorant Google Assistant pour ne plus à avoir à répéter le *wake-word*, Google se présente comme attentive aux évolutions des usages et aux problèmes rencontrés par ses clients. La firme simule vouloir faire attention à l'éducation des enfants ou rendre l'utilisation de l'enceinte plus simple pour les personnes âgées pour mettre en avant une vision de la technologie qui s'adapte à l'utilisateur et non l'inverse. En communiquant sur les évolutions de la Google Home, Google donne à l'utilisateur l'impression d'être écouté et d'avoir même un rôle à jouer dans l'évolution des enceintes connectées. Parallèlement, les GAFA restreignent leurs prises de parole sur la gestion des données ou sur la sécurité de leurs logiciels à moins qu'ils ne soient contraints de le faire. Le scandale de Cambridge Analytica, société qui avait récolté les données de 30 à 70 millions d'utilisateurs Facebook pour les analyser et utiliser dans le cadre de la campagne de l'ex-candidat Donald Trump a remis en lumière le débat sur la protection des données personnelles et a permis aux utilisateurs du

réseau social de pouvoir entendre Mark Zuckerberg sur le sujet au Congrès américain le 10 avril 2018.

Face à un manque d'initiative de la part des fabricants eux-mêmes pour protéger les données personnelles de leurs utilisateurs, les États interviennent. La seconde sous-partie évoque la réponse des instances publiques face aux stratégies de communication des marques qui éduquent les utilisateurs aux interfaces vocales et leur vendent une nouvelle vision, plus positive et rassurante de la technologie. En analysant les discours de la CNIL, nous avons compris pourquoi les enceintes connectées étaient au cœur du débat sur la protection des données personnelles. Les enceintes connectées, grâce à leurs fonctionnalités qui s'adaptent en fonction de l'utilisateur, peuvent être utilisées par des adultes et des enfants. Plus discrètes mais plus riches en connexions avec le quotidien des utilisateurs, elles enregistrent toutes les recherches Internet et les demandes de gestion domotique et collectent ainsi une masse de données considérable. Le taux d'adoption rapide de ces appareils alarme les instances publiques et on voit ainsi émerger un discours d'éducation préventive qui a pour objectif la sensibilisation du public à la collecte de ses données et à la violente incrustation de cette technologie dans son quotidien. Ce problème de société devient aujourd'hui un problème politique d'un point de vue européen dans la mesure où le continent réalise progressivement son manque d'initiative dans le secteur du numérique et observe une centralisation du pouvoir par les entreprises américaines. Comme nous l'avons vu avec le scandale Cambridge Analytica qui a mis en lumière le pouvoir d'influence de Facebook, les GAFAs participent aujourd'hui activement à l'évolution de la structure sociale et politique des sociétés. Pour assurer une bonne utilisation de ces objets numériques, la CNIL propose pour le moment de sensibiliser les utilisateurs à ce qu'ils font et à promouvoir d'autres solutions, françaises et plus respectueuses de la vie privée des citoyens.

## Conclusion

L'objectif de ce mémoire était de répondre à la problématique suivante : *Dans quelle mesure une enceinte connectée, alors qu'elle est censée être une interface naturelle et un support supplémentaire aux assistants vocaux déjà connus du public, nécessite-t-elle plusieurs leviers d'éducation tant sur son fonctionnement que sur sa place dans nos vies et nos sociétés ?*

Notre première hypothèse défendait l'idée qu'une éducation sur le fonctionnement des enceintes était nécessaire pour assurer une bonne utilisation aux usagers. La seconde hypothèse affirmait que les marques cherchent à transmettre une culture numérique à travers leurs publicités. La troisième hypothèse supposait, qu'en réponse aux discours séduisants des marques, les états réagissent et éduquent à leur tour.

Ce travail de recherche nous a ainsi permis de relever le double accompagnement que nécessitent les enceintes connectées. Le premier est en effet relatif à son fonctionnement. Malgré l'essence naturelle de la parole, les marques éduquent les utilisateurs à parler à leurs enceintes connectées. Si nous ne remettons plus notre langage courant en question, pratique que nous avons qualifiée d'infra-ordinaire, les fabricants d'enceinte connectée ont dû réaliser ce travail et réécrire les dialogues de notre vie quotidienne : « Quelle heure est-il ? » « Fera-t-il beau demain ? ». Cependant, les fabricants d'enceinte connectée n'ont pas encore réussi à programmer les assistants vocaux pour comprendre parfaitement notre langage. Dès lors, les utilisateurs sont face à une autre réalité et doivent s'adapter. Les enceintes connectées nous obligent à revoir notre manière de parler et à adapter nos habitudes orales au langage informatique. Ce travail d'adaptation a commencé pour certains usagers (des adultes déjà « connectés ») à la sortie des assistants vocaux ce qui leur a permis de prendre conscience des limites de la reconnaissance vocale et de l'imitation du langage humain. Les consommateurs ont réalisé qu'il était impossible de tout demander aux assistants vocaux et ont compris, à l'inverse, en quoi ils pouvaient leurs être utiles. Cependant, parce que le support est nouveau et surtout sans interface graphique, les fabricants doivent éduquer leur public à l'objet, même le plus technophile. Nous avons ainsi relevé que les marques profitaient de ce temps d'adaptation où l'utilisateur imite dans un premier temps ce qu'il voit autour de lui, et par conséquent dans les publicités, pour transmettre un langage propre à la consommation et créer certaines habitudes autour de l'interface vocale. La récurrence de certaines actions dans les publicités et les autres discours accompagnant les enceintes vocales nous prouvent que certaines fonctionnalités sont valorisées car créatrices de profit pour les fabricants – comme

par exemple l'usage des bibliothèques musicales qui sont souvent des services créés par les entreprises (comme Google Play Music ou Apple Music). D'autres fonctions, comme le minuteur ou la météo, sont fréquemment mises en scène parce qu'elles sont faciles à assimiler et créent donc un premier pas vers une utilisation quotidienne. Au-delà du langage, nous avons vu que le pari des fabricants était de combler le manque créé par l'absence d'interface graphique. Les utilisateurs ne sont pas habitués à un usage du numérique sans contact visuel. Ils sont habitués à une consommation très individuelle et personnelle des technologies permise par l'écran. Nous considérons en effet nos objets numériques comme intimes et personnels. L'utilisation de la voix fait de l'enceinte connectée un objet dont l'usage ne peut pas être totalement intime et reconditionne ainsi les rapports homme-technologie. L'enjeu des fabricants va au-delà d'une éducation au fonctionnement. Les fabricants doivent bouleverser les habitudes des consommateurs : changer leur relation à l'objet mais aussi créer une relation de confiance sans feedback visuel. En effet, l'unique feedback visuel, quand il y en a un, sont des petites lumières qui permettent - en général - de valider la prise en compte de la commande vocale. L'utilisateur n'est donc pas complètement perdu, sans aucun repère visuel, mais sa vue n'est que très peu sollicitée – un changement de taille dans ses habitudes numériques. L'objectif est de mettre l'utilisateur en confiance pour espérer qu'il passe à l'acte d'achat par simple commande vocale. Pour pallier ce manque et créer une relation de confiance, les entreprises diffusent différents messages qualifiant l'interface vocale comme étant la plus adaptée à nos quotidiens. En analysant les dispositifs de communication d'Amazon et Google, nous avons réalisé la volonté de créer une rupture avec la dépendance aux écrans dénoncée actuellement par la communauté scientifique. Amazon cible sa communication sur la valorisation d'une technologie accessible et complètement intégrée à nos quotidiens pour plus d'agilité. L'interface vocale apparaît comme un moyen de démocratiser la technologie, soulevant les barrières liées à la lecture et l'écriture infligées par l'interface graphique dont nous serions en plus, trop dépendants. Google positionne même son produit sur la déconnexion. L'enceinte connectée qui déconnecte, un oxymore révélateur du paradoxe de nos relations avec nos objets numériques. Pourtant, Google nous le dit : puisque nous sommes dépendants aux écrans, nous devons nous en séparer au moins pour les demandes ponctuelles et simples. Les publicités Google vont jusqu'à diaboliser l'écran qui apparaît comme une frontière entre son utilisateur et le monde extérieur. L'écran captif et isole alors que l'interface vocale permettrait d'être présent par le regard et d'insérer les usages numériques des utilisateurs plus facilement dans leur quotidien.

En 2018, les enceintes connectées étaient principalement utilisées pour faire des recherches sur Internet. Le sondage effectué par le cabinet PwC montre que pour les achats, 10% des personnes achètent quotidiennement avec une interface vocale, 40% le font souvent et 50% ne le font jamais. Les fabricants travaillent autant sur la routine des utilisateurs pour inverser la tendance et espérer que l'achat par commande vocale devienne prochainement un geste du quotidien.

Le développement de l'accès à internet et l'équipement des foyers ont permis aux enceintes connectées de se lancer au sein d'une population avertie et habituée à faire évoluer ses usages numériques. Cependant, au-delà de la technique, le problème auquel sont confrontées les marques est la méfiance du public. Pour assurer un bon développement des ventes des enceintes connectées, les fabricants ont misé sur une évolution de l'image de la technologie. Les publicités ont joué un rôle majeur dans cette transmission d'une vision rassurante et bienveillante des enceintes connectées. Les marques ont identifié puis contourné les peurs émergentes : la perte de contrôle de l'utilisateur sur les technologies et la surveillance de masse. Les marques rassurent sur la perte de contrôle liée à l'absence d'interface graphique en diabolisant l'écran et en ajoutant un contrôle physique avec l'ajout de différents boutons. L'écran devient alors l'interface qui prend le contrôle de l'utilisateur en le captivant à l'inverse de l'interface vocale qui libère l'usager. Pour ne pas apeurer les technophobes qui ne penseraient pas comprendre le fonctionnement de l'enceinte la qualifiant de trop éloignée de ce qu'ils connaissent, les marques n'ont pas misé – à part Apple – sur un design totalement lisse. Les enceintes sont majoritairement munies de boutons permettant à minima d'éteindre le micro ou de gérer le volume de l'appareil. La gestion physique du micro semble ainsi primordiale : elle donne l'impression à l'utilisateur de pouvoir facilement couper son enceinte, d'avoir le contrôle dessus. Sa présence rassure car elle rappelle à son usager qu'il peut cesser d'être écouté quand il le souhaite pour retrouver son intimité.

Au-delà d'une réflexion sur le design, les marques ont aussi réfléchi au rôle des enceintes connectées dans les foyers. Leurs publicités transmettent en effet une vision singulière de la place de la technologie dans la société. En analysant les dispositifs de communication d'Amazon, Google et Apple, nous avons relevé qu'il était récurrent que les marques donnent à leur enceinte le statut de gouvernante numérique. Au cœur de la maison, son rôle est en effet de gérer les emplois du temps de chacun, la domotique du foyer, d'aider aux devoirs et de servir d'aide-mémoire. Indispensable pour la famille, l'enceinte connectée devient le tableau de bord et aide-ménagère de la maison. Comme une gouvernante, elle agit aussi comme soutien moral. Au-delà de ses fonctions utilitaires, l'enceinte connectée participe

aussi aux « grands moments » de la vie de chacun. Les exemples les plus parlants sont les publicités d'Amazon et d'Apple qui valorisent l'enceinte en la présentant comme une amie, très proche de ses utilisateurs.

Les fabricants tiennent ainsi un double discours, l'un correspondant aux fantasmes « positifs » et l'autre aux fantasmes « négatifs » que l'imaginaire commun a sur les machines parlantes.

Les marques confortent en effet les utilisateurs dans ce qu'ils pensent être un assistant vocal intelligent : il assiste dans les tâches du quotidien et joue aussi le rôle de l'ami virtuel même si la fonction conversationnelle n'est que très peu valorisée. L'enceinte connectée ne peut pas nous parler « normalement » mais sa présence se veut rassurante.

Les marques développent également un discours qui répondrait aux fantasmes négatifs sur les enceintes connectées : la perte de contrôle sur la technologie et la surveillance de masse. Les fabricants ont en effet conçu une enceinte avec quelques boutons, laissant l'impression d'un contrôle sur la machine et ont ajouté le bouton *mute* qui permettrait à l'utilisateur d'être le commanditaire de sa connexion et du partage de ses données. Comme nous l'avons vu, l'objet permet pourtant peu de contrôle. L'application de gestion de l'enceinte connectée permet beaucoup plus de choix pour les paramètres et pourtant l'application liée au *smart speaker* n'est jamais évoquée. Les fabricants ne cherchent finalement pas à expliquer comment la machine fonctionne. Nous sommes loin du « the technology paradigm » que nous évoquions en première partie. L'objectif des marques n'est pas que l'utilisateur comprenne comment ses données sont enregistrées, où elles sont stockées ou comment elles sont utilisées pour affiner les profils publicitaires. Au contraire, cela pourrait nuire à l'image de marque. Alors les fabricants lissent les produits, les rendent simples et accessibles, justifiant « une démocratisation » de la technologie pour rendre la vie quotidienne des utilisateurs toujours plus facile. Mais est-ce leur rendre service que de les éloigner de la réalité technique des objets qu'ils utilisent ? Les marques tentent en effet de créer des habitudes liées aux Interfaces Utilisateurs Naturelles et ce sont ces habitudes qui sont à l'origine de la perte de contrôle des utilisateurs sur les technologies. Les consommateurs ne questionnent plus les objets qui les entourent et ne cherchent plus à comprendre comment ils fonctionnent. La technologie devient infra-ordinaire alors que la collecte de données personnelles est un sujet de société qu'il est important de questionner et d'améliorer.

En France, l'utilisation des données personnelles sont au cœur des préoccupations publiques depuis la fin du XXe siècle. Les enceintes connectées et les assistants vocaux ont

été l'objets de discours de prévention qui s'inscrivent plus généralement dans l'objectif de la CNIL : prévenir du contrôle des GAFAs sur le monde numérique. L'état voit en la discrétion des enceintes un danger pour les utilisateurs qui ne réalisent plus qu'ils donnent en continue une masse considérable d'informations sur leur vie privée. La CNIL devient un médiateur entre les consommateurs et les marques et tentent d'éduquer au mieux les citoyens aux risques liés au numérique. L'objectif de l'instance est de sensibiliser la population à ces dangers mais aussi de lui faire prendre conscience qu'un déséquilibre politique est en train de se créer entre les régions qui dominent le monde numérique et celles qui ne sont finalement que spectatrices. Mais au-delà d'un travail de sensibilisation au problème de souveraineté que soulèvent les GAFAs, les états doivent-ils repenser l'éducation technique de l'informatique ? La transparence entreprise-utilisateur ne peut-elle qu'exister qu'avec une population avertie, capable de comprendre le fonctionnement des machines ?

## **Bibliographie**

### Ouvrages

- Bachimont Bruno, « L'intelligence artificielle comme écriture dynamique : de la raison graphique à la raison computationnelle », Grasset, 1999
- Bachimont Bruno, « Le sens de la technique : Le numérique et le calcul », Encre marine, 2010
- Barthes Roland, « Mythologie », Point Essais, 2014
- Baudrillard Jean, « Le système des objets », Gallimard, 2014
- Bougnoux, Daniel, Sciences de l'Information et de la Communication. Larousse, Paris, 1993
- Cardon Dominique, « À quoi rêvent les algorithmes. Nos vies à l'heure des big data », Seuil, 2015
- Citton Yves, « L'économie de l'attention. Nouvel horizon du capitalisme ? », La découverte, 2014
- Donald Norman, « The Design of Everyday Things », 1988, New-York, Basic Books
- Dumouchel Paul et Damiano Luisa, « Vivre avec les robots. Essai sur l'empathie artificielle », Paris, Seuil, 2016
- Goffman Erving, « Les cadres de l'expérience », Le sens commun, 1991
- Jeanneret Yves, « Y'a-t-il (vraiment) des technologies de la communication ? », Presses universitaires du Septentrion, 2011
- Perec Georges, « L'infra-ordinaire », Editions du Seuil, Paris, Librairie du XXIe siècle, 1989
- Santolaria Nicolas, « Dis Siri. Enquête sur le génie à l'intérieur du smartphone », Anamosa, 2016
- Simondon Gilbert, « Du mode d'existence des objets techniques », Aubier, 2012

### Articles

- Cooper Alan, "The myth of metaphor", *Visual Basic Programmer's Journal*, 1955
- Ducheu David, « Effacer l'interface. Une trajectoire du design de l'interaction homme-machine », Interfaces numériques, Vol. 5 n°2/ juin-septembre 2016 - Designers interactifs (page 257 - 276)
- Katz Yarden, « Noam Chomsky on Where Artificial Intelligence Went Wrong », theatlantic.com, 1<sup>e</sup> novembre 2002

<https://www.theatlantic.com/technology/archive/2012/11/noam-chomsky-on-where-artificial-intelligence-went-wrong/261637/>

Le Marec Joelle, « Dialogue interdisciplinaire sur ‘l’interactivité’ », *Communication et Langages*, n°128, 2001

Myers Brad, « *A Brief History of Human Computer Interaction Technology* », *ACM interactions*. Vol. 5, no. 2, Mars, 1998

### Podcast

Podcast France Inter « Les assistants vocaux vont-ils faire de nos enfants des tyrans ? »  
4 juillet 2018, dans l’émission *C’est déjà demain* par Hélène Chevallier :  
<https://www.franceinter.fr/emissions/c-est-deja-demain/c-est-deja-demain-04-juillet-2018>

## **Annexes**

### **Entretien avec Yann Lechelle, COO de Snips**

Annexe 1 Page 79

### **L'illusion d'un contrôle sur les technologies – Le dashboard par Android, l'application Google Home, Blueprint par Amazon et la notice du Google Home**

Annexe 2 page 108

### **Extraits des études sur les usages des enceintes connectées**

Annexe 3 page 112

### **L'intégration des boutons sur les enceintes connectées**

Annexe 4 page 115

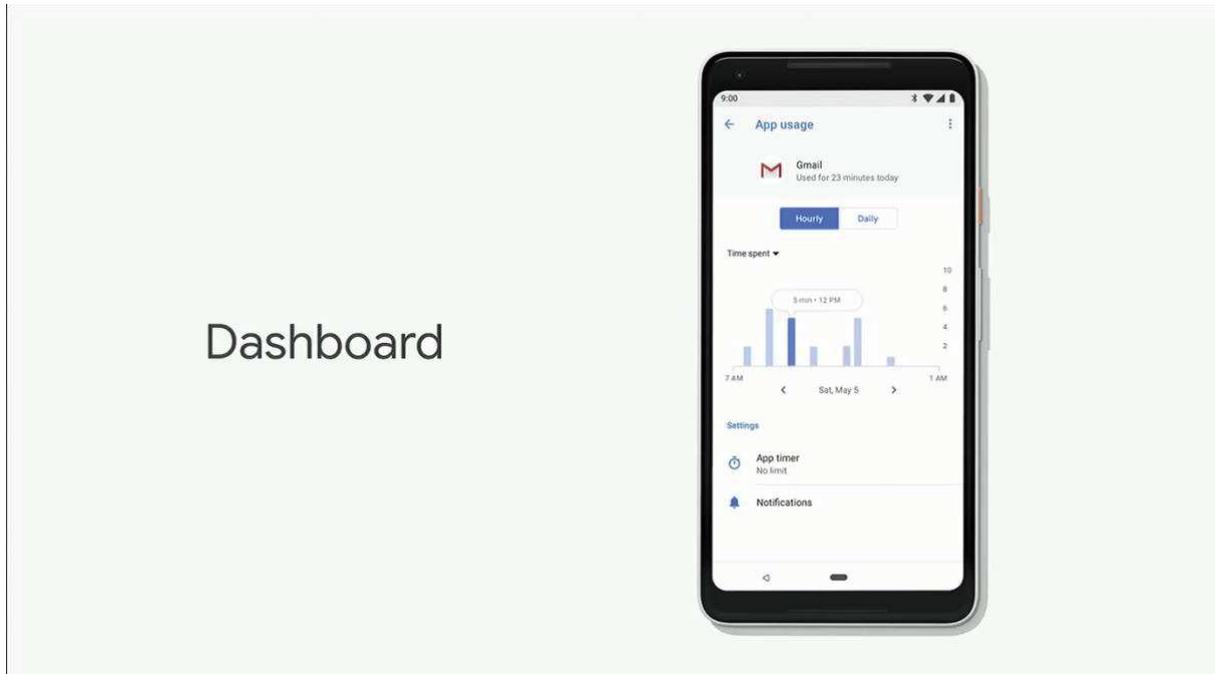
### **Tableaux d'analyse des discours de marque et analyse d'un corpus d'articles de presse**

Annexe 5 page 116

**Les pages 80 à 107 ont été retirées de la version diffusée en ligne.**

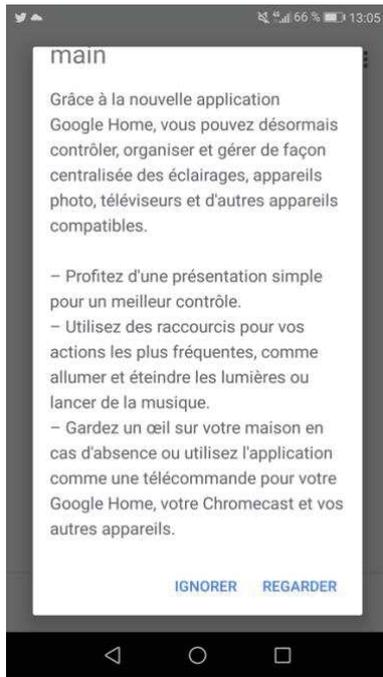
## Annexe 2

### L'illusion d'un contrôle sur les technologies – Le dashboard par Android, l'application Google Home, Blueprint par Amazon et la notice du Google Home



#### ***Présentation du dashboard pour Android (Google)***

*L'application dashboard d'Android permet à l'utilisateur d'avoir une vision générale sur son utilisation : il peut voir par exemple combien de temps il a passé sur son application Gmail et peut installer un « app timer » - une alerte qui lui communiquera qu'il a passé suffisamment de temps sur son application et qu'il faut arrêter. Il peut aussi gérer ses notifications. Dans l'ancienne version d'Android, on choisit si l'on veut ou pas recevoir des notifications d'une application. Le dashboard permet d'affiner ce choix et d'accepter de recevoir des notifications qu'à certains moments.*



### *Captures d'écran de l'application Google Home*

*« Profitez d'une présentation simple pour un meilleur contrôle. »*

*Google met en avant le contrôle sur la maison et les objets et fait du « contrôle » un argument de vente.*

[HOME](#)
[SKILLS YOU'VE MADE](#)
[HELP CENTER](#)

## Create personal Alexa skills in minutes

- 1 Pick a blueprint
- 2 Fill in the blanks
- 3 Use your skill



### Featured Blueprints

- 
**Custom Q&A**  
 Customize Alexa's response to your questions.
- 
**Fairy Tale**  
 Create an interactive prince and princess-themed tale.
- 
**How does it work?**  
 See how easy it is to create a skill with blueprints. [Watch Video](#)
- 
**Houseguest**  
 Welcome your guests with a guide to your home and neighborhood.
- 
**Quiz**  
 Challenge your friends with open-ended questions.
- 
**Doubles Trouble**  
 Play to find out: who couple knows each other best.

### More Games

- 
**Inspirations**
- 
**Burns**
- 
**First Letter**
- 
**Family Jokes**
- 
**Trivia**
- 
**Doubles Trouble**

### *Site de création de skills Alexa (Amazon)*

*Le site Blueprint permet à un utilisateur Alexa de créer ses propres skills. En choisissant parmi des templates, l'utilisateur peut créer un quiz, une blague, un jeu, ou simplement enregistrer une histoire ou certaines informations. Le site simplifie la création de skills, le but étant de permettre à n'importe qui de créer sa propre application vocale.*



Action	Appuyez sur l'appareil comme indiqué ci-dessous	Image
Lancer, mettre en pause ou arrêter l'alarme ou le minuteur 	Appuyez une fois la partie supérieure de l'appareil.	
Augmenter le volume 	Faites glisser le doigt dans le sens des aiguilles d'une montre sur la partie supérieure de l'appareil. <b>Remarque</b> : cela ne règle que le volume des médias, et non des alarmes et des minuteurs.	
Réduire le volume 	Faites glisser le doigt dans le sens inverse d'une aiguille d'une montre sur la partie supérieure de l'appareil. <b>Remarque</b> : cela ne règle que le volume des médias, et non des alarmes et des minuteurs.	
Lancer votre demande 	Appuyez longuement sur la partie supérieure de l'appareil.	
Désactiver le micro	Appuyez sur le bouton de désactivation du micro à l'arrière de l'appareil. <b>Remarque</b> : La désactivation du micro empêche Google Home d'écouter ou de répondre. Pour interagir avec Google Home, le micro doit être activé.	
Rétablir la configuration d'usine de l'appareil	Appuyez sur le bouton de désactivation du micro à l'arrière de l'appareil pendant environ 12 secondes.	

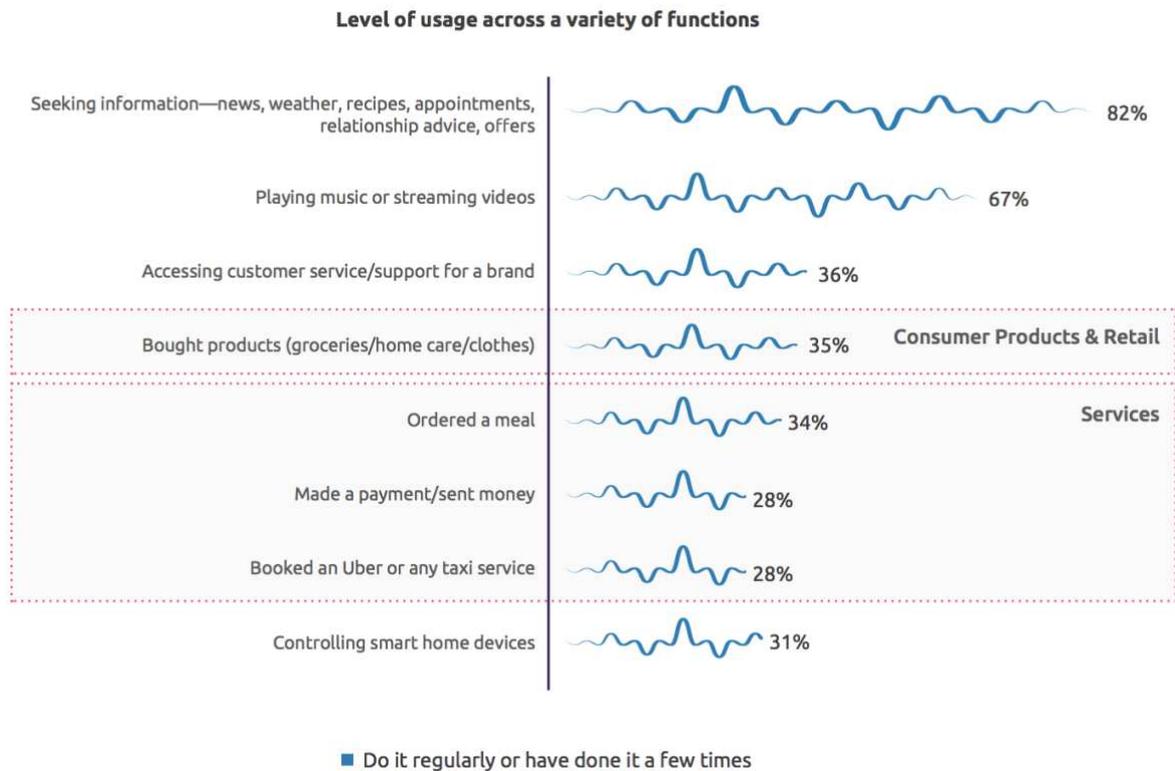
***Capture d'écran de la notice du « touch pad » sur le site du Google Home***

*Cette capture d'écran montre comment Google illustre le fonctionnement de son Google Home. On remarque que l'absence d'interface graphique nécessite un apprentissage de certains gestes, le touch pad étant simplement une zone tactile doté de led.*

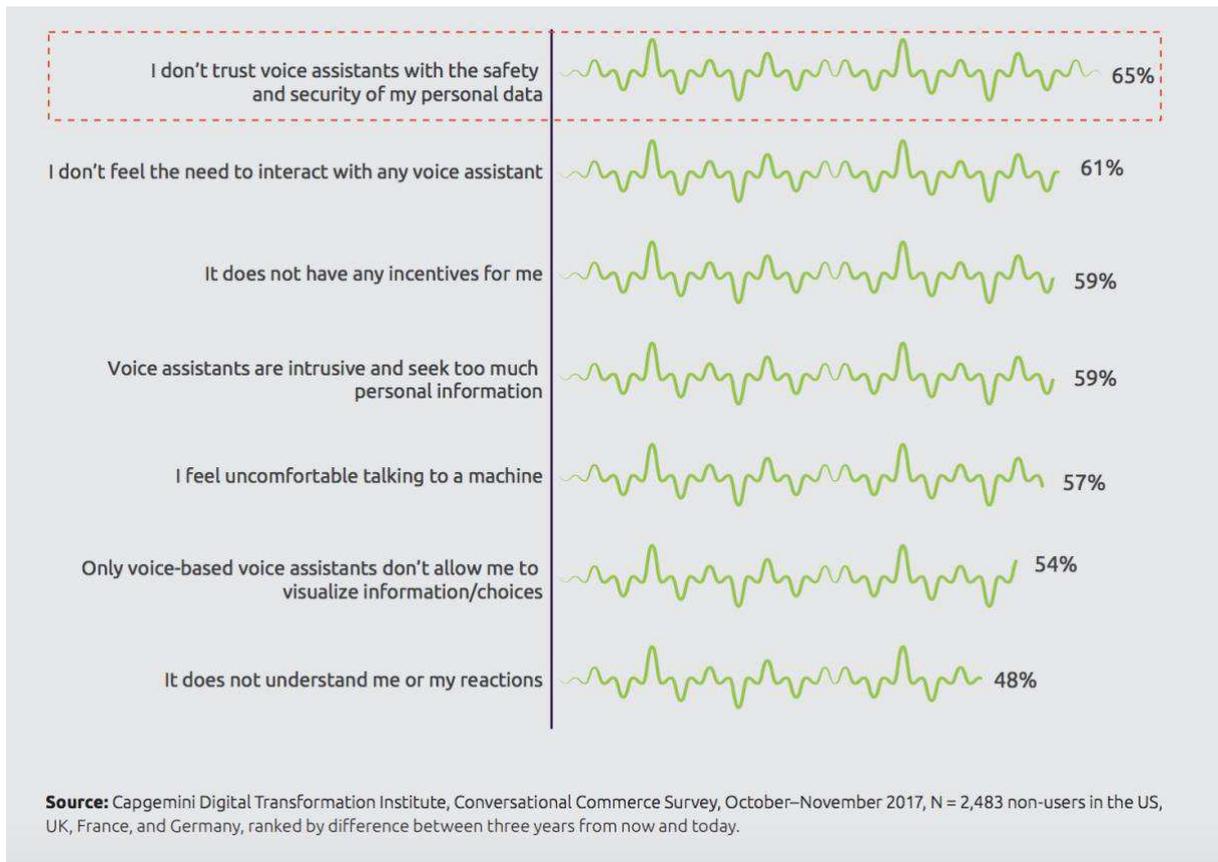
## Annexe 3

### Extraits des études sur les usages des enceintes connectées

Extraits de l'étude de Capgemini : Capgemini Digital Transformation Institute, Conversational Commerce Survey, October–November 2017, N = 2,558 users in the US, UK, France, and Germany.



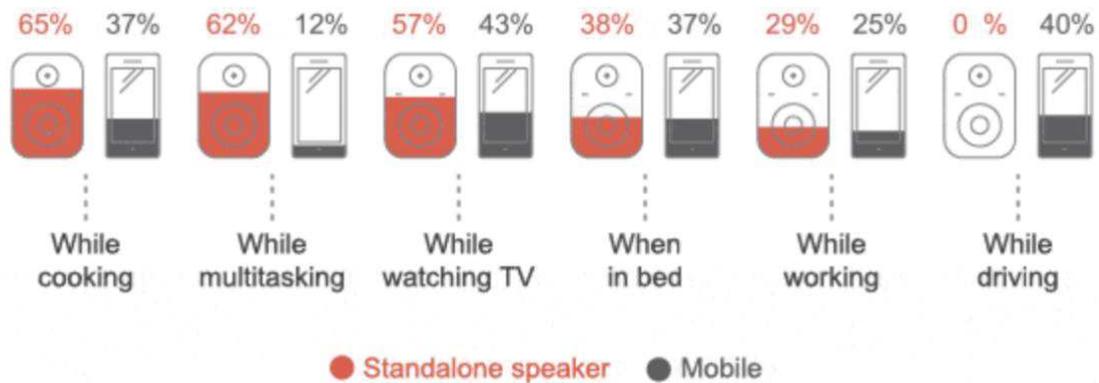
*L'étude met ici en évidence les usages des enceintes connectées : la demande d'informations est en première position alors que le contrôle des objets connectés et le paiement arrivent en dernières positions.*



*Ici, le sondage pointe le sentiment d'insécurité comme premier frein à l'utilisation d'une enceinte connectée. Alors que la majorité des interviewés utilisent les enceintes connectées pour demander la même chose qu'à leur moteur de recherche, ils considèrent que les enceintes connectées ne sont pas assez sécurisées.*

Extrait de l'étude de PwC - société d'audit et de conseils, 2018

Sondage réalisé avec un échantillon de 1000 Américains âgés de 18 à 64 ans qui ont accès à Internet.



Cette infographie nous montre que l'usage des enceintes connectées est concentré dans les moments où les mains sont prises : la cuisine, le multi-tâche. On remarque qu'il y a un fort taux d'utilisation lorsque les sondés regardent la télévision (la paresse prend le pas, mais surtout, les utilisateurs considèrent leurs enceintes assez intelligentes pour discerner leur voix des voix de la télévision). L'information intéressante est surtout celle qui considère les usages au lit : 38% des sondés utilisent plus leur enceinte que leur portable dans leur lit. Il aurait été intéressant de savoir quelle tâche les sondés demandent à leurs enceintes lorsqu'ils sont dans un lit – le smartphone étant un objet beaucoup plus intime et donc plus propice à entrer dans une chambre qu'une enceinte connectée.

## Annexe 4

### **L'intégration des boutons sur les enceintes connectées**

*Dans l'ordre, Echo d'Amazon, Tmall Genie d'Alibaba et Xiaomi Mi Smart Network Speaker de Xiaomi.*



## **Annexe 5**

### **Tableaux d'analyse des discours de marque et analyse d'un corpus d'articles de presse**

#### **Analyse des publicités**

Disponible avec le lien suivant :

<https://docs.google.com/spreadsheets/d/140Cq561kx8FErePz01hO4VRoDLEwBs8BAIh7SqbnRPs/edit?usp=sharing>

Extrait avec les publicités Google en page 117

#### **Analyse des articles de presse**

Disponible avec le lien suivant :

<https://docs.google.com/spreadsheets/d/1K0Is9NCjgsbAFc3GmEOvKjWfslhr6quQZQTqFwKS7c/edit?usp=sharing>

Extrait en page 118

Video	Date	Durée	Lien	Produit	Marché	Observations	Slogan	Moments représentés	Supports/médias présents	Personnages / Voix off	Ton de la voix / mod à utiliser	Placement et interaction avec le produit	Secateurs touchés	Réponse rassurante ? / ID des pairs
<b>GOOGLE</b>														
Introducing Google Home	04.10.2016	1'14	<a href="https://www.youtube.com/watch?v=0LLAVD0I-g">https://www.youtube.com/watch?v=0LLAVD0I-g</a>	Appareil Home	Appareil Home	Présentation générale du produit : Aucune présence humaine, juste le produit sur un fond blanc. Présentation : "Clean touch pad surface for total control" "Far-field voice recognition for hands-free use" "Mic mute button for privacy" "High excursion speaker for HiFi sound" "Modular design for customization"	Home by you. Hello by Google. Google Home.	La voix OFF montre comment appeler avec Google Home. "This is google Home, it is not a phone, it doesn't not have buttons or a screen. But you can use it to make phone calls [...] So you can set it to call your mom or ..."	Livres	Voix OFF, pas de personnage	Utilisation de l'impératif "Call X"	Au début sur un fond blanc puis dans différents environnements : sur une table de nuit, dans une cuisine, dans un salon. Mise en scène de Google Home dans différents contextes	Aucun ? Produit écrit Inconnu	> La peur de la perte de contrôle sans écran > Peur de ne pas être compris > Peur d'être toujours enregistré
Google home : hand free calling	26.08.2017	0'35	<a href="https://www.youtube.com/watch?v=9YR8BUd1CCQ">https://www.youtube.com/watch?v=9YR8BUd1CCQ</a>	Appareil Home	Appareil Home	Aucune présence humaine. Le produit est monté sur une table de salon, à côté d'une pile de livres et d'une lampe.	Phone calling on Google Home. Google Home	La voix OFF raconte "vous regardez vos amis à la cuisine et la voix de Google lui dit avec un ton désagréable) "Mais qu'est ce que c'est?"	Livres	Voix OFF, pas de personnage	Utilisation de l'impératif "Call X"	Produit placé au centre, sur une table de salon, à côté d'une pile de livres + lampe "Hey google, call this sandwich shop at west avenue"	Téléphone	La peur de la perte de contrôle sans écran
Google Home : callity			<a href="https://www.youtube.com/watch?v=ZyZCzICdLTC">https://www.youtube.com/watch?v=ZyZCzICdLTC</a>	Appareil Home	Appareil Home	Aucune présence humaine. Le produit est posé dans la cuisine et la voix raconte ce que le produit peut faire.		La voix OFF raconte à quel point on se sent important avec un assistant personnel parce qu'on peut lui donner des ordres.	Cuisine, bols	Voix OFF, pas de personnage	Utilisation de l'impératif "Put milk on my shopping order" "Put milk on my shopping list" "Turn off the light"	Produit placé au centre, sur un plan de travail de cuisine. A côté de bols et de vasos dans différents contextes : "Put milk on my shopping list" "Turn off the light"	Liste de course, positif / bouillon lumineux	
Google Home : show off	07.06.2017	0'35	<a href="https://www.youtube.com/watch?v=Z3D1N8BaMg">https://www.youtube.com/watch?v=Z3D1N8BaMg</a>	Appareil Home	Appareil Home	Aucune présence humaine. Le produit est posé dans la cuisine et la voix raconte ce que le produit peut faire.	"They want to like OMC's and they will be like 'I KNOW'!"	La voix OFF raconte "vous regardez vos amis à la cuisine et la voix de Google lui dit avec un ton désagréable) "Mais qu'est ce que c'est?"	La télévision	Aucun personnage mais dans l'histoire : un détenteur d'un compte de ses amis	Utilisation de l'impératif "Turn on the TV"	Au centre, dans une cuisine assez moderne, le produit est posé sur une table. On le découvre et un oignon et à ses côtés des bols et des condiments	La télévision, l'écran, l'ordinateur, l'assiette, l'assiette musicale	
Google Home - Profitez de vos vidéos à la voix	23.11.2017		<a href="https://www.youtube.com/watch?v=UjGd1YU1N8BaMg">https://www.youtube.com/watch?v=UjGd1YU1N8BaMg</a>	Appareil Home Mini	Appareil Home Mini	Google Home permet la reconnaissance avec l'instantané de la vidéo que l'entreprise et coupe du monde	Maintenant en petit format - Google Home Mini	Un père est sur sa tablette et ses deux filles jouent au jeu vidéo. Le père désigne la tablette et dit "Hey Google, joue la musique de ce jeu"	Tablette	Un père jeune / deux petites filles + une voix off masculine	Question + utilisation de l'impératif "OK google, quelle table est-ce que tu préfères?"	Au début pas de présence du produit, puis il apparaît sur une table basse. "VO" Et à quel point chose pouvait nous aider à être plus présent "Hey, Google, joue la musique de ce jeu"	Le smartphone, la tablette, l'ordinateur, l'assiette, l'assiette musicale	Pas de peur / ID rassurant, connu : (meuble IKEA dans l'environnement rassurant, connu)
Google Home - Godier de la semaine	28.09.2017		<a href="https://www.youtube.com/watch?v=UjGd1YU1N8BaMg">https://www.youtube.com/watch?v=UjGd1YU1N8BaMg</a>	Appareil Home	Appareil Home	Retour aux activités "simples" : le scrabble, les kapla, la graille, le dessin	Profitez du moment - Google Home	Un père avec sa petite fille et tous les amis de sa fille pour jouer à des jeux de société.	Livres	Père + petite fille + autres enfants	Utilisation de l'impératif "OK Google, met ma playlist hard rock"	La fille regarde sa fille qui a les yeux bandés à leur table et dit "Hey Google, met ma playlist hard rock"	L'écriteau musicale	Environnement rassurant, connu
Google Home - Wiscapou	27.09.2017		<a href="https://www.youtube.com/watch?v=YokkXZD0ic">https://www.youtube.com/watch?v=YokkXZD0ic</a>	Appareil Home	Appareil Home			Dans une (25 ans environ) jouant au scrabble, un soir une fille propose le mot Wiscapou - désaccord sur l'existence de ce mot.	Scrabble / bloc notes / crayon /	Amis, 25 ans environ, autour d'une table	Question, que veut dire Wiscapou?	Une des joueuses "OK Google, que veut dire Wiscapou?" GH est placé en arrière plan, sur une autre table derrière les joueurs - gros plan sur le produit à la fin, placé à côté d'un classeur.	Dictionnaire	Environnement rassurant, connu

Article	Source	Date	Auteur	Thèmes	Auteurs	Point de vue générale / Pronostic sur l'avenir du produit	Fonctionnalité de l'enceinte - ergonomie / éducation pratique	Education sociale - comportement à adopter selon les journalistes, les marques, les particuliers	Education politique - les paroles des institutions Jeu de pouvoir entre institutions et marques
On n'aura bientôt plus besoin d'écran... La révolution de la voix est en marche	L'ADN	22/2018	Olivier Robillart	Liste des nouvelles startups françaises qui se lancent dans la voix	La France / Les USA / Amazon, Google, Microsoft, Apple / Ava, Iily, Shlps, Xbrain, labo du CNRS et Thales	"Après le règne du pouce, outil indispensable de l'ère smartphone, voici celui de la voix." "Désormais, <b>la voix est capable de tout diriger.</b> " "Véritable tournant dans la technologie" Vision très optimiste : l'article présente les start-up FR qui se sont lancées dans l'économie de la voix et propose des services user-friendly - particulièrement pour ceux qui sont d'habitude oubliés des innovations : les personnes âgées et les personnes handicapées.	"On pourra bientôt tout faire avec la voix : Ecrire des mails en conduisant, rédiger et recifier des rapports, monter le chauffage ou fermer les volets, poser des questions à son assistant personnel, résoudre sa vie avec les services après-vente."	"Pour l'heure, le risque est essentiellement d'être submergé par des pubs. En effet, l'analyse des données conduit à la généralisation de publicités ciblées et géolocalisées. À terme, Home, démarches intègrent l'assistant vocal Alexa d'Amazon ou celui de la start-up française Shlps. Ces assistants vocaux sont essentiellement utilisés pour connaître la météo ou plier des objets connectés de la maison, comme des lampes."	"La Cnil alerte les consommateurs sur les risques inhérents à ce type d'objet, allant jusqu'à évoquer « la monétisation de l'intime »." « Chaque utilisateur doit ainsi intégrer que même si « la parole s'envole », ses requêtes vocales sont enregistrées dans le cloud, de la même manière qu'elles le seraient s'il les tapait au clavier dans certains moteurs de recherche », prévient la Cnil." "La Cnil recommande d'ailleurs de ne pas laisser les jeunes seuls dans une pièce avec ce type de dispositifs ou de les débrancher." "l'enceinte reste à l'écoute de tout ce qui se dit autour d'elle. Au point que la Cnil préconise de prévenir ses invités de la présence d'un tel appareil afin de les avertir que la conversation peut potentiellement être enregistrée. Ambrance : "La Cnil recommande d'éviter de relier ses serrures, volets ou alarmes connectés à un assistant vocal." "Vider l'historique des données collectées, partir du tableau de bord de ces ap- pareils."
Les enceintes connectées écoulent tout à la maison	Le Figaro	12/5/2017	Elisa Bembaron	Danger des enceintes connectées	La Cnil / les consommateurs / Google, Amazon, Shlps / les enfants / Burger King / Netatmo / Symantec	"Les enceintes connectées s'imposent sur les listes de Noël" L'article se base sur les recommandations de la Cnil, et de l'étude de Semantec pour guider les consommateurs dans leur utilisation des enceintes : Ne pas laisser les enfants utiliser une enceinte seule Ne pas lier son enceinte aux serrures, alarmes, etc Supprimer régulièrement ses historiques Prévenir ses invités du potentiel enregistrement de leur propos	"Concrètement, ces enceintes sont capables de comprendre les questions et les ordres formulés à haute voix et de les appliquer. Qu'ils s'agissent de Google Home, démarches intègrent l'assistant vocal Alexa d'Amazon ou celui de la start-up française Shlps. Ces assistants vocaux sont essentiellement utilisés pour connaître la météo ou plier des objets connectés de la maison, comme des lampes."	"Pour l'heure, le risque est essentiellement d'être submergé par des pubs. En effet, l'analyse des données conduit à la généralisation de publicités ciblées et géolocalisées. À terme, Home, démarches intègrent l'assistant vocal Alexa d'Amazon ou celui de la start-up française Shlps. Ces assistants vocaux sont essentiellement utilisés pour connaître la météo ou plier des objets connectés de la maison, comme des lampes."	"La Cnil alerte les consommateurs sur les risques inhérents à ce type d'objet, allant jusqu'à évoquer « la monétisation de l'intime »." « Chaque utilisateur doit ainsi intégrer que même si « la parole s'envole », ses requêtes vocales sont enregistrées dans le cloud, de la même manière qu'elles le seraient s'il les tapait au clavier dans certains moteurs de recherche », prévient la Cnil." "La Cnil recommande d'ailleurs de ne pas laisser les jeunes seuls dans une pièce avec ce type de dispositifs ou de les débrancher." "l'enceinte reste à l'écoute de tout ce qui se dit autour d'elle. Au point que la Cnil préconise de prévenir ses invités de la présence d'un tel appareil afin de les avertir que la conversation peut potentiellement être enregistrée. Ambrance : "La Cnil recommande d'éviter de relier ses serrures, volets ou alarmes connectés à un assistant vocal." "Vider l'historique des données collectées, partir du tableau de bord de ces ap- pareils."