COLLECTION **LIVRE BLANC** **N°1**

À VOTRE ÉCOUTE

Exploration des enjeux éthiques, techniques et juridiques des assistants vocaux

CNIL.
COMMISSION NATIONALE
INFORMATIQUE & LIBERTÉS

COLLECTION **LIVRE BLANC - N°1**

À VOTRE ÉCOUTE
Exploration des enjeux éthiques, techniques et juridiques des assistants vocaux

SOMMAIRE

Septembre 2020

Directeur de la publication : Gwendal Le Grand

Rédacteur en chef : Bertrand Pailhès

Rédacteurs de ce livre blanc : Martin Biéri et Félicien Vallet avec l'aide d'Isabelle Corbara, Pauline Faget, Basile Guley, Estelle Hary et Juliette Hirtz.

Conception graphique : Agence Linéal - 03 20 41 40 76

Impression : DILA

Dépôt légal : à publication

Illustrations : Malte Müller

Les points de vue exprimés dans cette publication ne reflètent pas nécessairement la position de la CNIL.

Sont vivement remerciés Monir Azraoui, Nacéra Bekhat, Régis Chatellier, Clément Commerçon, Antoine Courmont, Eric Delisle, Hugo Dussert, Louis Dutheillet de Lamothe, Pauline Girard, Jérôme Gorin, Armand Heslot, Mathias Moulin, Aymeric Pontvianne, Flora Sanchez et Clémence Scottez ainsi que l'ARCEP, le CNPEN, le CSA et la Hadopi.

ÉDITO

Devenus incontournables depuis quelques années, les assistants vocaux sont en passe d'ouvrir une nouvelle ère d'utilisation des outils numériques qui nous entourent, de notre téléphone à notre téléviseur, en passant par notre véhicule ou nos équipements électroménagers. Une ère qui serait conversationnelle et interactive et qui pourrait révolutionner nos manières de communiquer, de consommer, d'accéder à l'information… et donc de vivre.

En permettant une fluidification des échanges, une simplification des commandes et en offrant une facilité d'utilisation, de tels dispositifs pourraient susciter un réel progrès en matière d'inclusion numérique et sociale. Ils pourraient s'inscrire de façon durable dans le contexte quotidien de leurs utilisateurs, notamment pour les personnes en situation de dépendance, qu'elles soient âgées ou handicapées.

Ces avancées indéniables ne doivent cependant pas occulter les questions que les assistants vocaux posent du point de vue de la protection des données, notamment de la transparence du fonctionnement de leur système.

En premier lieu, la matière première qu'ils manipulent est très fortement ancrée dans notre intimité. Premier outil de communication avec nos semblables, la voix est en effet partie prenante de notre identité et révèle, outre le sens des mots, de nombreuses informations sur l'émetteur d'un message oral. Âge, genre, condition physique, accent, origine géographique et socio-culturelle, éducation, état de santé ou émotionnel, mais également identité – la voix étant une caractéristique biométrique permettant l'identification – sont autant d'informations qui peuvent être déduites du signal audio. L'actualité a illustré les risques liés à l'écoute accidentelle de la sphère intime et témoigne de l'indispensable confiance que l'on doit attendre des constructeurs de ces objets, mais également de ceux qui les déploient, comme de ceux qui les alimentent, en créant des applications par exemple.

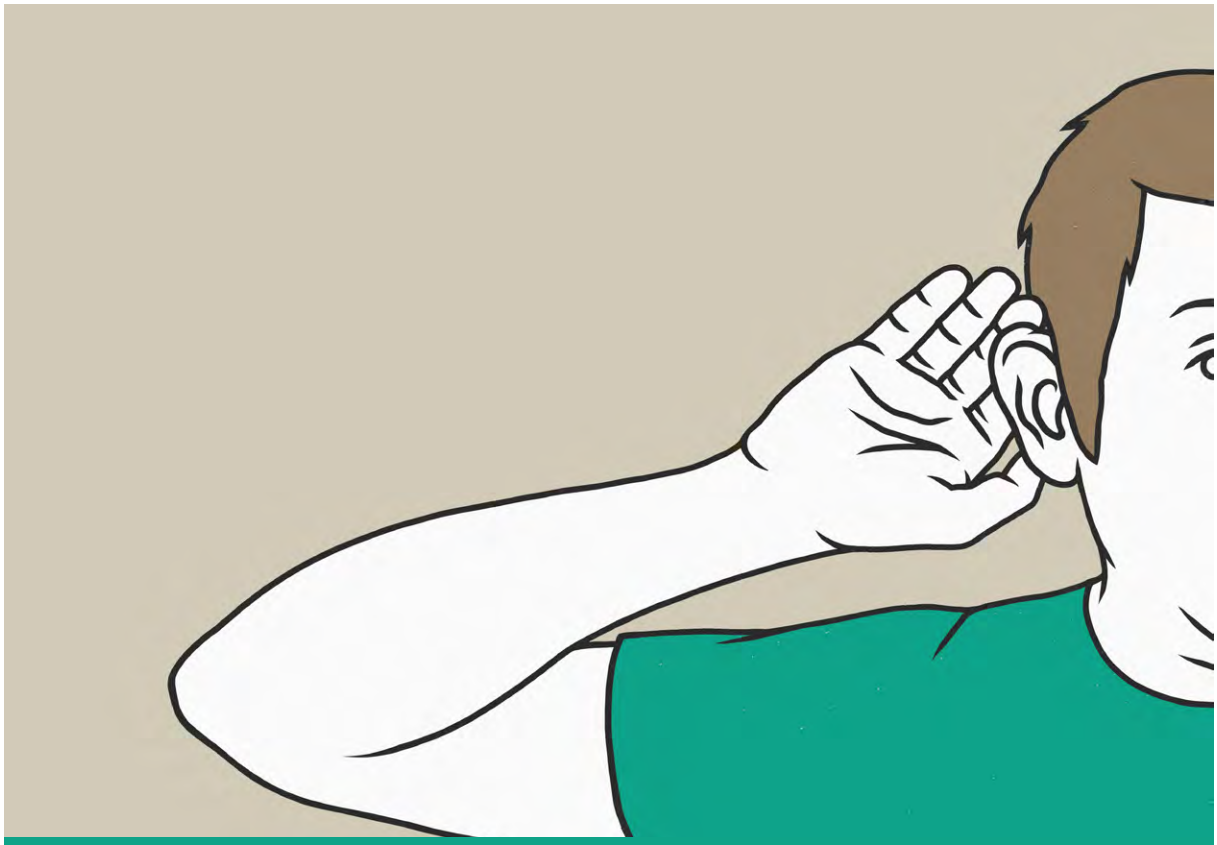
En second lieu, si, selon l'adage commercial largement répandu, les assistants vocaux réussissent à faire « disparaître la technologie », l'usage qui est fait des données ne doit en aucun cas être rendu invisible. Sous leur aspect pratique et ludique, la réalité de certains de ces objets est de capter nos habitudes de vie pour enrichir un profil, ce qui peut contribuer à nous enfermer dans un écosystème marchand. Ces nouveaux avantages compensent-ils leurs inconvénients ? Comment, dès lors, les encadrer pour qu'ils soient respectueux de la vie privée des individus ?

Ces questions, la CNIL les explore depuis plusieurs années et sous différents angles. Depuis 2017 et l'arrivée dans nos foyers des assistants vocaux intégrés dans des enceintes de salon, la CNIL a publié de nombreux articles et entretiens avec des experts sur son site institutionnel (cnil.fr) ou sur celui du LINC, le Laboratoire d'Innovation Numérique de la CNIL (linc.cnil.fr), dédié aux études, expérimentations et analyses prospectives. La CNIL a suivi le développement de ces dispositifs en étant au contact des différents concepteurs et accompagne, à travers de nombreux partenariats tels que celui noué avec Inria (l'Institut national de recherche en sciences et technologies du numérique), les travaux de recherche qui sont menés sur le sujet.

L'objectif de ce livre blanc est ainsi de restituer ces travaux de manière accessible et à tout type de public. Il s'agit de présenter les différents enjeux juridiques, techniques ou éthiques, et de répondre aux préoccupations de ceux qui construisent ces assistants, de ceux qui les déploient, comme de ceux qui les utilisent. Enfin, il vise également à proposer des conseils et des orientations pour contribuer à un développement d'outils respectueux des droits fondamentaux des individus les utilisant.

Puisse ce livre blanc proposer quelques pistes de réflexion sur l'usage des données par les assistants vocaux, notamment le respect des grandes principes prônés par le RGPD, et contribuer à une réflexion collective indispensable à ce nouveau rapport au numérique.

Marie-Laure Denis, *Présidente de la CNIL*



ASSISTANTS VOCAUX, DE QUOI PARLE-T-ON ?

S'il est possible d'enregistrer le son - et donc la voix - depuis près d'un siècle et demi, la parole demeure encore aujourd'hui associée à une certaine volatilité. Toutefois, la généralisation des usages des technologies de traitement automatique de la parole et leur intégration dans un nombre croissant d'objets dessine un rapport nouveau à « l'objet vocal ». Un changement de paradigme essentiel pour les utilisateurs semble ainsi à prévoir.

LA SPÉCIFICITÉ DE LA VOIX

Parler, échanger, communiquer, dire, énoncer, raconter, converser, etc. Les termes sont nombreux pour caractériser les échanges oraux que nous entretenons quotidiennement. Pour autant, notre voix n'en demeure pas moins une grande inconnue.

Une donnée intime

La voix est un élément essentiel de la construction de l'identité. Elle est une réalisation concrète du langage, qui est lui-même une capacité de communication. C'est notamment cette faculté langagière parlée et complexe qui nous distingue des autres espèces animales. En pratique, la voix véhicule en dehors du discours (les mots proprement dits) des caractéristiques nombreuses de l'individu : émotions, intentions, condition physique, etc. En se reposant sur des mécanismes de perception, nos interlocuteurs sont en mesure d'interpréter ces signaux et de décrypter ces états.

On distingue généralement, dans la communication, des activités de « haut niveau », intégrant des processus intellectuels (perception, conception du discours, décodage et analyse du message, etc.) et de « bas niveau », permettant la réalisation concrète de l'échange (mécanismes physiques de l'articulation et de production des sons, de l'audition, etc.). On peut donc considérer la voix comme un processus de « bas niveau » puisqu'il s'agit d'un outil physique au service de la pensée et du discours. En 1916, Ferdinand de Saussure introduit la notion de signe linguistique qui illustre cet état de fait[1]. Celui-ci unit « non pas un nom et une chose, mais un concept et une image acoustique », l'image acoustique étant appelée signifiant et le concept qu'il désigne signifié.

Ainsi, notre voix est porteuse de messages plus complexes que le simple « sens » des mots prononcés (le signifié). Ces informations non-verbales contiennent les éléments dits paralinguistiques : silences, intonations, gestes, posture du corps, ton de la voix, expressions faciales, etc. Le signal vocal permet donc l'extraction d'informations nombreuses et diverses : la signification du message bien évidemment, mais également des indications relatives à l'âge, au sexe, à la condition physique, à la familiarité avec la langue, à l'accent, à l'origine géographique et socio-culturelle, à l'éducation, à l'état de santé ou émotionnel, etc.

Une question à…
Joana Revis

Tout au long de notre vie, notre voix nous accompagne. Quelle(s) relation(s) entretenons-nous avec elle ?

Notre voix fait tellement partie de nous que c'est la toute première chose que nous faisons à la naissance : pousser un cri ! Elle nous accompagne effectivement tout au long de notre vie, elle change au fil du temps, elle évolue, elle est là, tout le temps sans que nous y pensions et la plupart d'entre nous entretenons finalement une relation assez ingrate avec elle : elle est là et c'est la moindre des choses, et nous n'en prenons pas vraiment soin. Il n'y a finalement que deux situations dans lesquelles nous en prenons conscience : lorsque nous entretenons un rapport passionnel avec elle (c'est le cas des chanteurs par exemple ou des comédiens), ou lorsque nous la perdons (au cours d'une simple laryngite ou dans le cas de lésions chroniques des cordes vocales). Là, tout d'un coup, nous nous rendons compte de son importance.

Joana Revis est orthophoniste-vocologiste et maître de conférences associée à la faculté de médecine Aix Marseille Université

> Entretien intégral à retrouver sur LINC
Joana Revis, *Notre voix porte en elle toutes les intentions qui sont les nôtres*, Linc.cnil.fr, mars 2018, https://linc.cnil.fr/fr/joana-revis-notre-voix-porte-en-elle-toutes-les-intentions-qui-sont-les-notres

1 - Ferdinand de Saussure, *Cours de linguistique générale*, Payot, 1916

Vue comme un simple élément du langage non-verbal, la voix est souvent peu considérée pour ce qu'elle est. Pour Joana Revis, au contraire, celle-ci occupe une place centrale dans notre société basée sur la communication[2]. Notre voix est un miroir de l'âme, capable notamment de chanter, rire, exiger, énoncer, convaincre, pleurer, rassurer, reprocher, implorer, consoler, prévenir, manipuler, jouer et ainsi exprimer ce qui est indicible, trahir des émotions, ou encore caractériser une personnalité. La voix n'est jamais univoque. Elle varie au cours du temps pour un même individu tout en restant profondément singulière. Il est normal dans ces conditions que nous développions des relations très fortes avec elle et avec celles de ceux qui nous entourent.

Une donnée volatile

Une autre des grandes spécificités de la voix tient à son caractère à la fois intangible et volatile. Physiquement, la voix n'est qu'une trace laissée par des mouvements d'air causés par le phénomène de phonation, c'est-à-dire la production de sons propres à la langue parlée. Toutefois, depuis l'invention du phonographe par Thomas Edison en 1877, il est possible d'inscrire ces traces sur des enregistrements et par la suite de les rejouer et de les analyser. Avec les évolutions de la technique, la voix est devenue très facilement captable, et cela, potentiellement à l'insu même des personnes[3].

Chaque individu bénéficie de droits de la personnalité qui ont pour but de le protéger. Parmi ces droits, le droit à la protection de la vie privée et le droit à l'image sont reconnus comme deux droits subjectifs, distincts, qui tendent à la protection de l'intégrité morale. Si le premier a été affirmé en 1948 par la Déclaration universelle des droits de l'homme des Nations unies (article 12) et en droit français en 1970 dans l'article 9 du Code civil, c'est au XIX^{e} siècle, de façon contemporaine à l'invention de l'enregistrement sonore (et de la photographie) que le droit à l'image a été reconnu. Lorsqu'on traite du droit à l'image en général, il est fréquent de l'associer à l'image visuelle. Toutefois, comme cela a été précisé par le tribunal de grande instance de Paris le 19 mai 1982 dans la cadre d'un procès initié par la cantatrice Maria Callas suite à la radiodiffusion non autorisée d'enregistrements de travail, « la voix est un attribut de la personnalité, une sorte d'image sonore »[4]. En écho au droit à l'image d'une personne, il convient donc également de prendre en compte le droit à sa voix.

L'article 226-1 du Code pénal dispose qu'« est puni d'un an d'emprisonnement et de 45 000 euros d'amende le fait, au moyen d'un procédé quelconque, volontairement de porter atteinte à l'intimité de la vie privée d'autrui en captant, enregistrant ou transmettant, sans le consentement de leur auteur, des paroles prononcées à titre privé ou confidentiel ». Cette disposition protège de façon englobante toutes les paroles prononcées dans un cadre privé, les images appartenant à une personne ou la représentant, les informations relatives à son domicile ou aux lieux qu'elle fréquente, les informations relatives à son état de santé, ses courriers et ses courriers électroniques privés, les informations relatives à sa vie familiale ou à sa vie amoureuse ou encore ses opinions politiques, religieuses ou philosophiques. Différents éléments de jurisprudence existent et permettent de caractériser cette atteinte. En pratique, on s'intéresse souvent au cumul de différents critères à savoir : le caractère clandestin de l'enregistrement, sa localisation, sa durée, etc.

La voix n'est jamais univoque. Elle varie au cours du temps pour un même individu tout en restant profondément singulière.

2 - Joana Revis, *La voix et soi : Ce que notre voix dit de nous*, DeBoeck, 2013.
3 - Félicien Vallet, *Les droits de la voix (1/2) : Quelle écoute pour nos systèmes ?*, Linc.cnil.fr, mai 2019, https://linc.cnil.fr/fr/les-droits-de-la-voix-12-quelle-ecoute-pour-nos-systemes
4 - TGI Paris, 19 mai 1982, aff. Maria Callas, Recueil Dalloz 1983, jurisprudence p. 183 ou encore Paris, 22 janv. 2001, Dalloz 2002, p. 2375, note A. Lepage - Adde, D. Huet-Weiller, La protection juridique de la voix humaine, RTD civ. 1982, p. 497

Une question à...
Nicolas Obin

De plus en plus de sociétés proposent des produits permettant de créer un clone numérique de sa voix ou de celle d'un tiers. Quels sont les usages commerciaux possibles ?

La possibilité de reproduire la voix d'une personne de manière naturelle et réaliste a ouvert de nombreuses applications possibles dans lesquelles les entreprises se sont engouffrées. Les voix de synthèse sont amenées à nous accompagner au quotidien que ce soit avec les smartphones, les assistants à la maison ou embarqués dans les véhicules, les agents d'accueil virtuels et les centrales d'appel automatisées, pour ne donner que quelques exemples. Des entreprises proposent aujourd'hui de pouvoir créer une voix de synthèse avec sa propre voix, au prix d'une certaine quantité d'enregistrements à réaliser. Des essais préliminaires ont été réalisés dès 2014 pour pouvoir parler dans une autre langue mais en conservant sa propre voix, avec comme application directe la traduction automatique. À terme, il sera sans doute possible de créer des avatars vocaux pour les joueurs de jeux vidéo, pour créer la voix des personnages, ou pour réaliser le doublage automatique d'un acteur dans des langues différentes. Le clonage de la voix en synthèse vocale a également des applications médicales importantes, comme la « prothèse vocale ». Toutes ces possibilités, si elles sont fascinantes et annoncent le devenir de ces technologies, n'en sont néanmoins qu'à un état encore largement embryonnaire.

Nicolas Obin est maître de conférences au laboratoire des Sciences et Technologies de la Musique et du Son (STMS), Ircam - CNRS - Sorbonne Université.

> Entretien intégral à retrouver sur LINC
Nicolas Obin, *La voix artificielle rend la machine plus humaine*, Linc.cnil.fr, mars 2019, https://linc.cnil.fr/fr/nicolas-obin-la-voix-artificielle-rend-la-machine-plus-humaine

Une donnée reproductible

Un autre aspect du droit à la voix concerne « l'identité vocale ». Si l'acte de falsification de contenus audiovisuels n'est pas nouveau, les récentes possibilités offertes par la synthèse de parole renouvellent les questions relatives à la fraude et à l'usurpation d'identité[5]. En particulier, l'application de technologies d'intelligence artificielle dites de *deepfake* – mot-valise formé à partir des termes anglais *deep learning* (apprentissage profond) et de *fake* (faux) – exploitent de puissantes techniques d'apprentissage automatique (*machine learning*) pour manipuler ou générer des contenus visuels et sonores à fort potentiel de tromperie. En pratique, les principales méthodes utilisées aujourd'hui sont les auto-encodeurs et les réseaux adverses génératifs (*generative adversarial networks* ou GANs)[6]. Si aujourd'hui la qualité des clones vocaux produits à partir de faibles quantités d'échantillons audio ne semble pas encore satisfaisante[7], le sujet, qui tenait il y a encore peu du domaine de la science-fiction, concentre de plus en plus l'attention. Du point de vue du droit, alors que les cas d'usurpation vocale numérique sont encore très rares[8], voire inexistants, la « confusion d'apparence » est sanctionnée par la justice depuis près de cinquante ans dans le cadre d'imitations vocales. En 1975, un jugement a sanctionné l'utilisation d'un spot publicitaire télévisé reposant sur un texte lu par une personne dont « la diction, le débit, le ton et les inflexions de voix [...] évoquaient les particularités verbales du comédien Claude Piéplu (la voix des Shadoks) »[9]. À l'étranger, et en particulier aux États-Unis, la jurisprudence fait état de cas similaires, par exemple pour les chanteurs Tom Waits et Bette Midler[10]. Toutefois, l'imitation d'une personne peut dans certains cas être autorisée, notamment si celle-ci est justifiée par le contexte historique ou d'actualité dans lequel est située l'œuvre, si elle n'est pas de caractère diffamatoire ou présente un caractère de parodie ou de caricature.

Une donnée à géométrie variable

En plus de nous assurer lors de nos échanges de la compréhension du sens du message que nous recevons, nous analysons la manière dont celui-ci nous est délivré. Cette caractérisation multiniveaux révèle ainsi des informations bien plus riches qu'une simple suite de mots juxtaposés ; elle est donc, à ce titre, une donnée à manipuler avec précaution.

5 - Félicien Vallet, *Les droits de la voix (2/2) - Quelle parole pour nos systèmes ?*, Linc.cnil.fr, juin 2019, https://linc.cnil.fr/fr/les-droits-de-la-voix-22-quelle-parole-pour-nos-systemes
6 - Ruben Tolosana et al., *DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection*, arXiv, janvier 2020, https://arxiv.org/pdf/2001.00179.pdf
7 - Jaime Lorenzo-Trueba et al., *Can we steal your vocal identity from the Internet?: Initial investigation of cloning Obama's voice using GAN, WaveNet and low-quality found data, Odyssey*, 2018, https://arxiv.org/pdf/1803.00860.pdf
8 - Morgan Tual, *« Deepfake » : dupée par une voix synthétique, une entreprise se fait dérober 220 000 euros*, Le Monde, septembre 2019, https://www.lemonde.fr/pixels/article/2019/09/06/deepfake-dupee-par-une-voix-synthetique-une-entreprise-se-fait-derober-220-000-euros_5507365_4408996.html
9 - Affaire Claude Piéplu, TGI Paris 3 décembre 1975, D. 1977, p. 211, note R. Lindon.
10 - Daniel Payette, *Les autres facettes de l'image: le nom, la voix et la ressemblance*, Les Cahiers de la propriété intellectuelle, 2015, https://www.lescpi.ca/s/162

Le Règlement général sur la protection des données (RGPD) dispose que toute information relative à une personne physique susceptible d'être identifiée, directement ou indirectement, est une donnée à caractère personnel. Une telle définition recouvre de nombreux types de données : nom, numéro d'immatriculation, numéro de téléphone, photographie, date de naissance, commune de résidence, empreinte digitale, historique de géolocalisation, etc. Dans le cas qui nous intéresse, des informations dérivées du signal vocal émis par un individu peuvent être également données à caractère personnel. Véhicule privilégié de nos interactions sociales, l'analyse que nous réalisons de la voix peut permettre la compréhension du message transmis, l'identification de son émetteur, mais également la catégorisation de celui-ci selon différentes modalités. Il s'agit donc d'une donnée personnelle qui, en fonction de l'utilisation qui en est faite, est à géométrie variable. Il convient de s'assurer que les principes cardinaux de la protection des données – qui concernent notamment la pertinence du traitement de données, sa transparence, le respect des droits des personnes, la maîtrise des données ou encore la gestion des risques et la sécurité – sont bien respectés.

L'utilisation la plus évidente des données vocales s'avère être la transcription textuelle des mots et phrases prononcés qui permet de « décoder » le sens du message transmis. En pratique, il s'agit de faire correspondre aux mouvements d'air captés par un microphone, la suite de mots prononcée par la personne. Dans son article 9, le RGPD prévoit une interdiction de principe au traitement des données « sensibles » tout en ménageant certaines exceptions telles que l'obtention du consentement des personnes concernées par exemple. La transcription automatique de la parole peut ainsi faire apparaître certaines de ces données « sensibles » comme des informations relatives aux opinions politiques, aux convictions religieuses ou philosophiques, à l'appartenance syndicale ou encore à la vie sexuelle. D'autres informations susceptibles de se trouver dans des enregistrements audio peuvent également être considérées comme des données sensibles et ne sont pas nécessairement relatives aux seuls mots prononcés. Ainsi, divers traitements automatiques peuvent concerner l'origine raciale ou ethnique[11]. Des données relatives à la santé peuvent également être inférées. À titre d'exemple, des recherches sont menées depuis plusieurs années pour caractériser la présence de maladies dégénératives comme celles d'Alzheimer ou de Parkinson[12]. En effet, parmi les manifestations cliniques de celles-ci, on observe que les troubles de la voix surviennent à un stade précoce de développement de la maladie.

Enfin, parmi les données sensibles au sens du RGPD, on trouve également les données biométriques, c'est-à-dire les données permettant ou confirmant l'identification d'un individu par ses caractéristiques physiques, physiologiques ou comportementales, lorsque leur traitement vise à identifier une personne de façon unique. Elles sont produites par le corps lui-même et le caractérisent de façon définitive. Elles peuvent parfois être utilisées pour suivre et identifier un individu, même à son insu. Ces données sont particulièrement sensibles car permanentes. Qui plus est, elles peuvent dans de nombreux cas être captées à

Une question à…
Jean-François Bonastre

Quelles sont les évolutions des technologies vocales à venir ?

La voix porte beaucoup d'informations sur l'individu comme son âge, son sexe, ses origines, son éducation, ses ressentis, son état physique ou psychique voire même ses intentions… Bien entendu, ou peut-être heureusement, nous ne savons pas décrypter toutes ces informations avec une fiabilité suffisante pour une exploitation réelle. Pas encore du moins… Rechercher si la consommation d'alcool ou de stupéfiant peut se voir dans la voix est une piste que plusieurs laboratoires explorent. Enfin, des travaux et des applications possibles tournent autour de la détection des émotions ou des attitudes émotives. Certains vont même jusqu'à l'évaluation de la sincérité, soit un détecteur de mensonge qui ne dit pas son nom… Même si la rigueur scientifique et les résultats ne sont pas toujours présents, des sessions sur ces sujets sont régulièrement proposées lors des grandes conférences scientifiques du domaine, souvent autour de « challenges » mettant en compétition des systèmes et donnant l'impression que tout est résolu…

Jean-Francois Bonastre est professeur au Laboratoire d'Informatique d'Avignon et spécialiste du traitement de la parole et de l'authentification vocale

> Entretien intégral à retrouver sur LINC
Jean-François Bonastre, *La voix n'est pas une biométrie classique,* Linc.cnil.fr, février 2017, https://linc.cnil.fr/fr/jean-francois-bonastre-la-voix-nest-pas-une-biometrie-classique

11 - Sonja Trent-Brown, *Voice quality: Speaker identification across age, gender, and ethnicity,* The Journal of the Acoustical Society of America, mars 2018
12 - Laetitia Jeancolas et al., *L'analyse de la voix comme outil de diagnostic précoce de la maladie de Parkinson : état de l'art, Compressions et Représentation des Signaux Audiovisuel,* mai 2016

Une question à…
Chloé Clavel

En pratique, comment encode-t-on les émotions et s'assure-t-on de leur correspondance avec un phénomène physique ?

En quoi consiste le phénomène émotionnel ? La réponse à cette question est un sujet de controverse. Doter la machine des capacités de compréhension des comportements humains : tel est le défi scientifique autour duquel se rassemblent différentes communautés scientifiques (traitement du signal, traitement automatique du langage, intelligence artificielle, robotique, interaction homme-machine, etc.). Les informations disponibles sont les signaux acquis par le système via des capteurs (image, son, capteurs physiologiques). Les données manipulées sont donc de très bas niveau : les échantillons sonores ou encore les pixels des images. Dans l'exemple de la voix, une grande partie des descripteurs acoustiques utilisés pour caractériser les différents états émotionnels est destinée à modéliser les modifications du signal acoustique liées à des modifications physiologiques à la base de la glotte. Les modifications corporelles ou physiologiques qui accompagnent certains états émotionnels, vont fortement influer sur le mode de production du message oral du locuteur. Par exemple, dans le cas de la peur, les modifications physiologiques typiques sont l'augmentation du pouls et de la pression du sang et la sécheresse de la bouche, et se manifestent par une voix plus forte, et plus aigüe et un débit plus rapide, au contraire de l'ennui et de la tristesse qui sont corrélés avec un abaissement du rythme cardiaque et se manifestent par une voix plus grave, moins intense et un débit plus lent.

Chloé Clavel est professeure associée en Affective Computing à Telecom Paris

> Entretien intégral à retrouver sur LINC
Chloé Clavel, *Les machines ne font « pas encore » mieux que les humains pour interpréter les émotions*, Linc.cnil.fr, octobre 2018, https://linc.cnil.fr/itw-chloe-clavel-les-machines-ne-font-pas-encore-mieux-que-les-humains-pour-interpreter-les-emotions

distance, sans que la personne ne le sache, notamment pour la voix ou l'image. Appliquée au cas de la voix, la mise en œuvre d'un système ayant pour objectif de reconnaître un individu à partir de ses caractéristiques vocales – on parle de reconnaissance du locuteur – est un traitement de données biométriques. Au cours des années, la CNIL a développé une doctrine relative à l'encadrement de l'utilisation des données biométriques, qu'il s'agisse de reconnaissance d'empreinte digitale, de visage, d'iris, de silhouette… ou encore de reconnaissance du locuteur[13]. Plusieurs expérimentations ont par exemple été autorisées par la CNIL pour l'authentification vocale d'usagers de banques de détail sur les serveurs vocaux interactifs[14]. Pour cela, des mesures techniques et organisationnelles ont notamment dû être mises en œuvre afin de satisfaire les impératifs de protection des données et en particulier pour garantir aux personnes les utilisant la maîtrise de leurs données biométriques.

La voix peut également permettre d'inférer des informations très intimes sans pour autant que celles-ci soient considérées comme sensibles par la réglementation. À titre d'exemple, de plus en plus de sociétés proposent d'analyser le signal vocal afin d'en extraire des informations relatives à l'état émotionnel d'un individu. Les objectifs avancés sont nombreux : permettre à des téléconseillers de connaître et analyser l'humeur de leurs interlocuteurs en temps réel, s'assurer de leur professionnalisme et de la bonne tenue de poste, analyser les postures, attitudes et savoir-être de candidats ayant enregistré lettres de motivations et CV vidéo[15], autant d'applications dont la mise en œuvre pose de nombreuses questions[16]. Ainsi, si de telles données ne disposent pas d'un statut particulier dans la réglementation en matière de protection des données, il n'en résulte pas moins que leur exploitation est susceptible de procurer un sentiment d'intrusion aux personnes concernées[17].

13 - CNIL, Biométrie, https://www.cnil.fr/fr/biometrie
14 - CNIL, La CNIL autorise l'expérimentation de dispositifs biométriques de reconnaissance vocale par des établissements bancaires, https://www.cnil.fr/fr/la-cnil-autorise-lexperimentation-de-dispositifs-biometriques-de-reconnaissance-vocale-par-des
15 - Anne Rodier, *Le robot, fidèle compagnon du recruteur,* Le Monde, février 2020, https://www.lemonde.fr/economie/article/2020/02/22/le-robot-fidele-compagnon-du-recruteur_6030459_3234.html
16 - Laurence Devillers, Les robots émotionnels: santé, surveillance, sexualité… et l'éthique dans tout cela ?, L'observatoire, 2020.
17 - Régis Chatellier, *Captation des émotions : comment vous le direz pourra être retenu contre vous…*, Linc.cnil.fr, avril 2018, https://linc.cnil.fr/fr/captation-des-emotions-comment-vous-le-direz-pourra-etre-retenu-contre-vous

ASSISTANT VOCAL, QUI ES-TU ?

Si les questions relatives au traitement de la voix ne sont pas nouvelles, l'arrivée progressive d'assistants vocaux dans nos équipements personnels les remet en perspective. Pour bien comprendre les enjeux soulevés par ces nouveaux dispositifs, il est indispensable de comprendre d'où ils viennent et comment ils fonctionnent.

Historique des assistants vocaux et de leurs ancêtres

1769 — Wolfgang von Kempelen
Première machine mécanique permettant la synthèse vocale

1879 — Thomas Edison / Charles Cros
Invention du phonographe

1922 — Radio Rex (Elmwood Button Co.)
Premier jouet activable par la voix (en disant «Rex»)

1933 — Ernest Esclangon (Observatoire de Paris)
Invention de l'horloge parlante, premier dispositif d'automatisation de la diffusion de l'heure par téléphone

1939 — Voder (Bell Labs)
Premier synthétiseur de parole électronique

1952 — Audrey (Bell Labs)
Machine capable de transcrire les chiffres de 0 à 9

1961 — Shoebox (IBM)
Machine capable de transcrire 16 mots et les chiffres de 0 à 9

1964 — Eliza (MIT)
Programme de simulation d'un psychothérapeute sous forme textuelle

1976 — Harpy (Carnegie Mellon)
Machine capable transcrire plus de 1000 mots (financement DARPA) sous forme textuelle

1987 — Poupée Julie (Worlds of Wonders)
Jouet que les enfants peuvent entraîner à répondre à leur voix

1990 — Dragon Dictate (Dragon)
Premier système de transcription automatique grand public

1996 — Clippy (Microsoft)
Outil d'accessibilité destiné à aider les utilisateurs de la suite Office

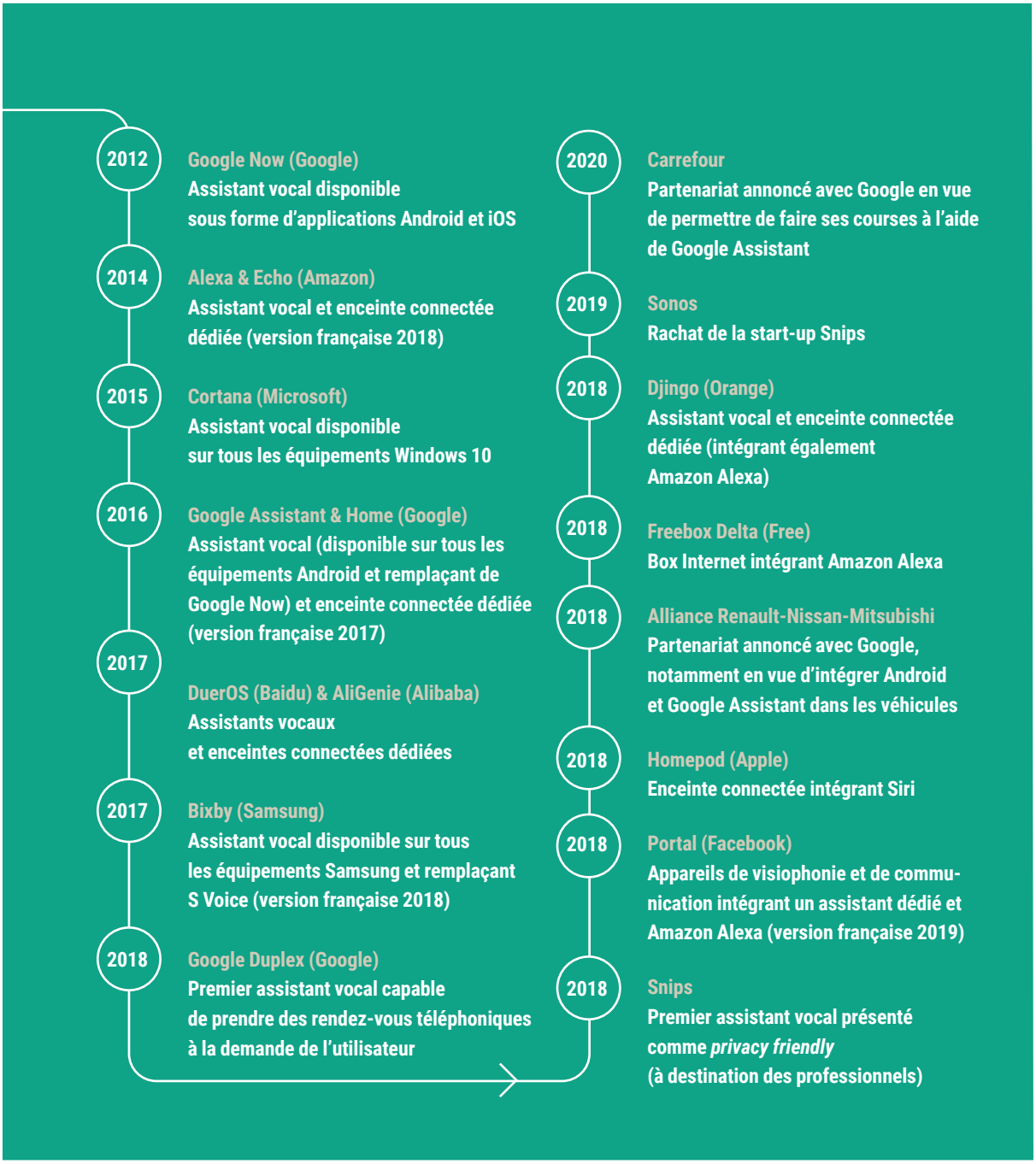
2002 — Office (Microsoft)
Intégration systématique d'outils de transcription automatique dans la suite Office

2011 — Watson (IBM)
Programme de traitement du langage naturel, champion du jeu télévisé américain Jeopardy

2011 — Siri (Apple)
Premier assistant vocal mis à disposition du grand public sur iOS (version française 2012)

Une histoire déjà ancienne...

Ce n'est que depuis les années 2010 que les assistants vocaux ont fait irruption pleinement dans l'imaginaire collectif. Pourtant, de nombreuses étapes ont précédé leur entrée en scène[18]. En effet, pas d'assistant vocal sans capacités d'enregistrement numérique des sons, de transcription automatique de la parole et de compréhension du langage naturel ou encore de synthèse vocale...



18 - Wolfgang Minker et Françoise Néel, *Développement des technologies vocales*, Le travail humain, 2002, https://www.cairn.info/revue-le-travail-humain-2002-3-page-261.htm

ZOOM SUR...

Machines parlantes et fiction

Que ce soit dans la littérature ou au cinéma, le rôle de l'assistant parlant a revêtu différentes facettes. Deux grands types semblent se détacher : la fiction a ainsi permis d'en faire tour à tour un partenaire de l'homme, ou au contraire, un adversaire à l'intelligence froide, responsable de prises de décisions radicales contre l'humain, lui par nature irrationnel. Dans le premier cas, on retrouve une machine parlante qui incarne un rôle de sympathique acolyte (*side-kick*) jusqu'à celui de garde-fou : c'est le cas notamment de J.A.R.V.I.S., qui permet à Iron Man de ne pas mourir dans chaque film (Jon Favreau, 2008 et suivants), de GERTY dans le film *Moon* (Duncan Jones, 2009) ou encore de la voiture KITT dans la série *K2000* (Glen A. Larson, 1982-1986). À l'opposé, les exemples d'assistants calculateurs, froids et sans pitié ne manquent pas non plus. Le plus célèbre d'entre eux est bien évidemment HAL 9000, dans *2001, l'Odyssée de l'espace* (Stanley Kubrick, 1968) mais le Master Control Program de *Tron* (Steven Lisberger, 1982) ou V.I.K.I. de *I, Robot* (Alex Proyas, 2004) peuvent également être cités. Un peu à part, la question de la relation homme-machine est interrogée de manière plus individuelle et sociale comme par exemple dans le film *Her* (Spike Jonze, 2013), dans lequel l'assistant Samantha fait tourner la tête de Joaquin Phoenix. La machine devient alors un idéal amoureux, une partenaire parfaite ou une amie intime comme illustré dans la série *Mr. Robot* (Sam Esmail, 2015), où l'agent du FBI Dominique DiPierro partage ses pensées les plus personnelles avec son assistant vocal Alexa.

Les fictions amènent également des réflexions sur la notion d'habitat du futur, et sur les rôles d'assistants domestiques. La série *Years and Years* (Russell T. Davies, 2019) montre la place prise par les assistants dans les prochaines années, en en faisant l'un des principaux artisans des échanges dans la famille. Ce dispositif introduit de nouvelles façons de communiquer, de s'appeler, d'envoyer des messages, etc. Dans une version plus ancienne, c'est un côté plus dystopique que l'on peut retrouver dans *Retour vers le futur 2* (Robert Zemeckis, 1986), film dans lequel la maison connectée est la seule à répondre à Marty McFly (Michael J. Fox) lorsque celui-ci rentre du travail, l'accueillant par des surnoms comme « seigneur du manoir » ou « roi du château ». Enfin, on pourrait remonter jusqu'à Blanche-Neige (Jacob et Wilhelm Grimm, 1812) et le miroir magique de la sorcière, qui répond à ses interrogations après qu'ont été prononcés les mots « Miroir, miroir ».

D'autres œuvres et ouvrages peuvent également être cités, sans pour autant que cette liste soit exhaustive : La zone du dehors (Alain Damasio, 1999), WOPR dans *WarGames* (John Badham, 1983), Data dans *Star Trek* (Gene Roddenberry, 1966), *Blade Runner* (Ridley Scott, 1982), Alpha 60 dans *Alphaville* (Jean-Luc Godard, 1965), Icarus dans *Sunshine* (Danny Boyle, 2007), Auto dans *Wall-E* (Andrew Stanton, 2008), etc.

Un assistant vocal, c'est quoi ?

On distingue les langages naturels des langages formels. Un langage formel dispose d'une sémantique qui est fixée pour être non-ambiguë, en vue de développer un programme informatique. Un opérateur humain qui développe ou souhaite interagir avec un programme informatique développé en langage formel doit donc s'y adapter, tant en termes de lexique que de syntaxe. À contrario, le langage naturel dispose d'une sémantique propre au langage humain. Suivant les caractéristiques de la langue et la diversité du lexique, une même instruction peut être formulée de multiples manières alors que certaines commandes peuvent sembler similaires mais se rapportent à deux objets différents. On a alors fréquemment recours à des mécanismes d'inférence afin de lever ces ambiguïtés, par exemple, suivant ce qui a été dit précédemment,

« Un assistant vocal peut être défini comme une application logicielle offrant des capacités de dialogue oral avec un utilisateur en langage naturel. »

l'heure où l'instruction a été énoncée, le lieu, les centres d'intérêt de la personne, etc.

Un assistant vocal peut-être décomposé en modules permettant de réaliser différentes tâches : captation et restitution des sons, transcription automatique de la parole (*speech to text*), traitement automatique de la langue, stratégies de dialogue, accès à des ontologies (ensembles de données et concepts structurés relatif à un domaine donné) et sources de connaissances externes, génération de langage, synthèse vocale (*text to speech*), etc. Concrètement, l'assistant doit permettre l'interaction afin de réaliser des actions (« allume la radio », « éteins la lumière ») ou d'accéder à des connaissances (« quel temps va-t-il faire demain ? », « le train de 7 h 43 circule-t-il ? »). Il joue ainsi un rôle d'intermédiaire et d'orchestrateur censé faciliter la réalisation de tâches de l'utilisateur.

En pratique :

Un assistant vocal n'est pas une enceinte intelligente... mais une enceinte intelligente peut être équipée d'un assistant vocal

Il est fréquent de confondre assistant vocal et enceinte intelligente. Toutefois la seconde n'est qu'une incarnation matérielle (*form factor* en anglais) du premier. Un assistant vocal peut être déployé dans un smartphone, une enceinte intelligente, une montre connectée, un véhicule, un équipement ménager, etc.

Un assistant vocal est un assistant personnel... mais la réciproque n'est pas nécessairement vraie

Un assistant personnel est un agent logiciel visant à interagir en langage naturel avec un individu afin de l'aider à réaliser des tâches. Toutefois, la voix n'est pas nécessairement la modalité d'interaction utilisée. Il peut ainsi s'agir d'échanges textuels, comme par exemple dans le cas d'un *chatbot*.

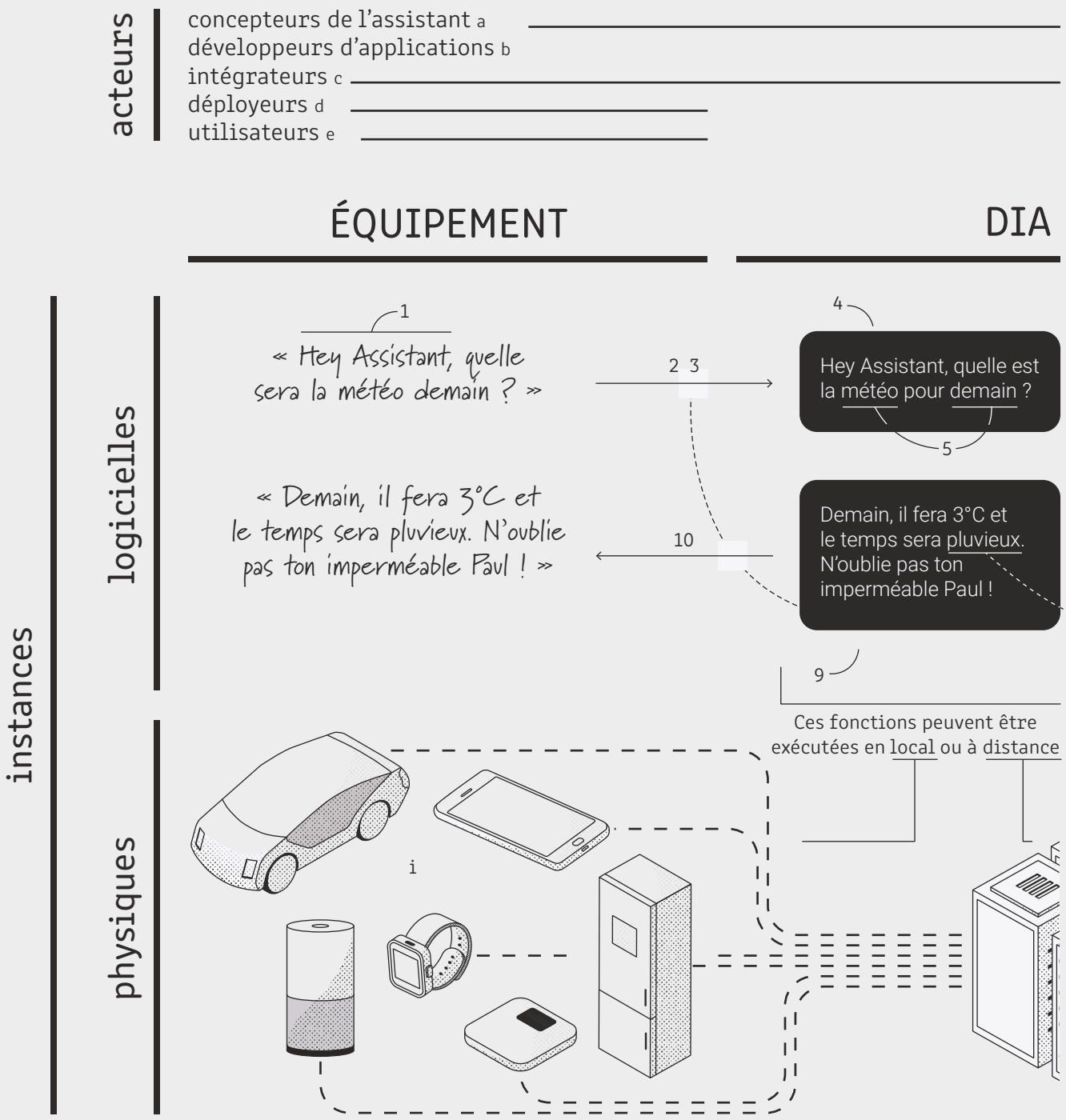
Si l'interaction avec l'assistant vocal se matérialise par l'échange oral d'un utilisateur avec un équipement électronique, en pratique l'organisation du traitement de données sous-jacent peut mettre en jeu de multiples schémas de circulation de l'information.

Il est possible d'isoler trois grandes entités pour en comprendre le fonctionnement (voir infographie page 14) :

• **l'instance physique :** élément matériel dans lequel s'incarne l'assistant (smartphone, enceinte, réfrigérateur, etc.) et qui embarque des microphones, haut-parleurs et capacités de calcul (plus ou moins développées en fonction des cas) ;

• **l'instance logicielle :** partie mettant en œuvre l'interaction homme-machine à proprement parler et qui intègre les modules de transcription automatique de la parole, de compréhension et génération du langage naturel, de dialogue et de synthèse vocale. Celle-ci peut être opérée directement à l'intérieur de l'élément matériel, mais est dans de très nombreux cas réalisée de manière distante ;

• **les ressources :** données externes telles que les bases de connaissances, ontologies ou applications métiers qui fournissent la connaissance (« indiquer l'heure sur la côte ouest des États-Unis ») ou permettent de réaliser concrètement l'action demandée (« augmenter la température de 1,5 °C »).

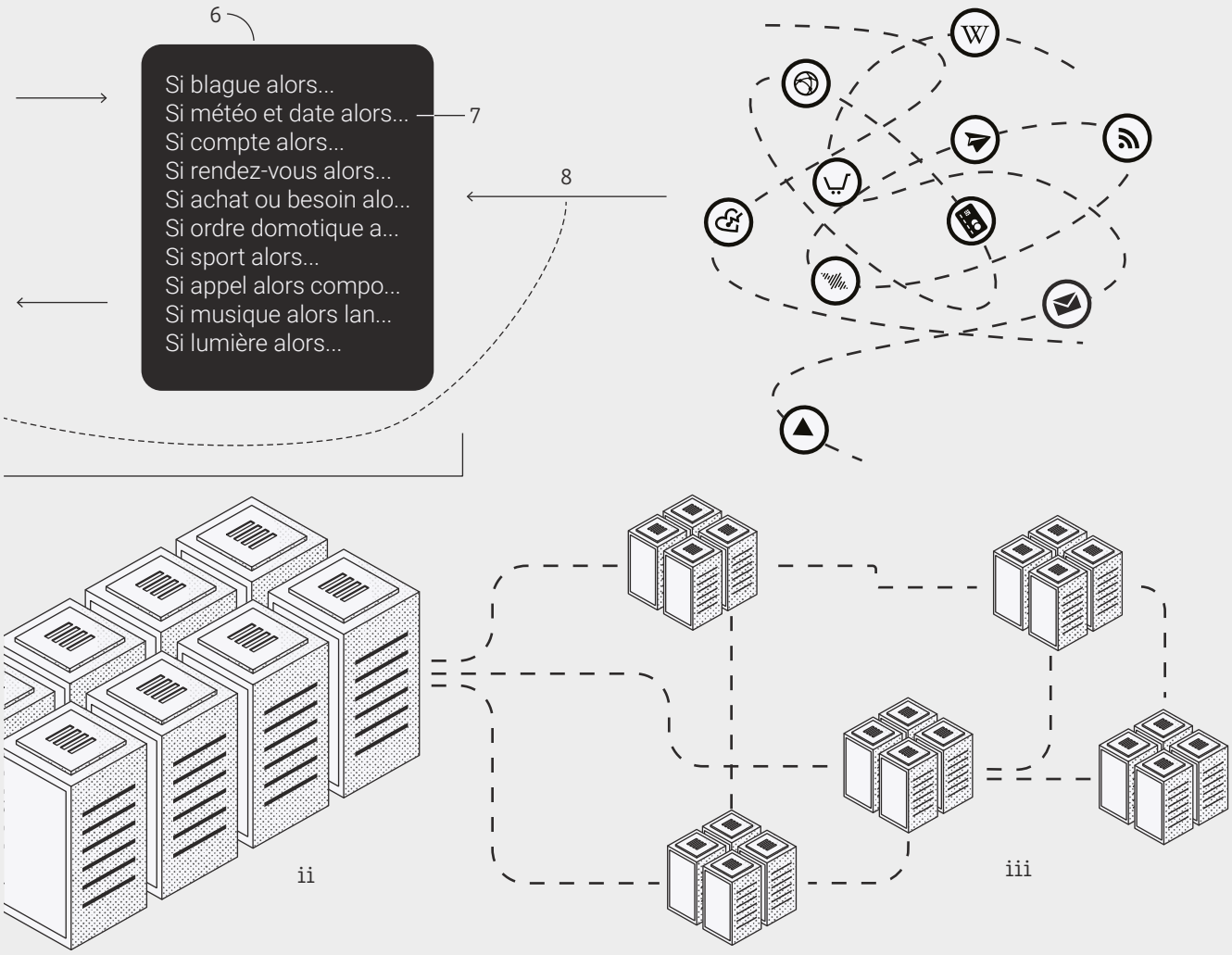


LA VOIE DES DONNÉES

Une infographie pour comprendre le fonctionnement des assistants vocaux.

LOGUE

RESSOURCES



légendes

Acteurs

a. Développe les modalités de fonctionnement de l'assistant.
b. Développe des applications à déployer sur l'assistant.
c. Intègre l'assistant dans ses objets.
d. Met à disposition des assistants dans des espaces sous sa responsabilité.
e. Utilise un objet embarquant un assistant vocal.

Instances logicielles

1. Détection locale du mot clé
2. Vérification du mot clé
3. Reconnaissance du locuteur
4. Transcription automatique *(speech-to-text)*
5. Détection d'intentions
6. Gestionnaire de dialogue
7. Sélection d'intentions
8. Informations récupérées sur des ressources publiques ou accessibles par authentification
9. Génération de langage naturel
10. Synthèse vocale *(text-to-speech)*

Instances physiques

i. Objets domestiques de consommation
ii. Serveurs des concepteurs de l'assistant
iii. Serveurs des développeurs d'application

Comme détaillé dans l'infographie page 14 et 15, une succession de tâches s'opère pour la réalisation d'une action ou l'accès à une information.

0) Déployé au sein d'un équipement (smartphone, enceinte, véhicule), l'assistant vocal est en veille. Concrètement, il est en permanence à l'écoute, tout en concentrant éventuellement son écoute sur certaines zones de l'espace, par exemple pour essayer de neutraliser des sources sonores tierces comme un téléviseur (à l'aide la technique de filtrage spatial ou *beam forming*). Mais, il ne conserve pas les données audio et ne procède à aucune opération tant qu'un mot-clé spécifique n'a pas été entendu. Pour cela une mémoire tampon de quelques secondes est utilisée.

1) L'utilisateur prononce un mot clé et « réveille » l'assistant. Un canal d'écoute s'ouvre et le contenu audio est transmis à la volée (*stream*).

2) Dans bien des cas, si le traitement est réalisé de façon distante, une seconde vérification de la prononciation du mot-clé est faite côté serveur afin de limiter les déclenchements intempestifs.

3) Éventuellement, et s'il a été préalablement enrôlé – c'est-à-dire si un apprentissage de ses caractéristiques vocales a été réalisé à partir d'échantillons de voix qu'il aura produits – le locuteur peut être identifié (*speaker identification*).

4) L'utilisateur énonce sa requête et celle-ci est transmise aux instances de traitement. Il peut s'agir de serveurs distants, ou dans le cas d'un traitement local, de ressources matérielles embarquées dans l'équipement. La séquence de parole prononcée est alors automatiquement transcrite (*speech to text*).

5) À l'aide de technologie de traitement automatique du langage naturel (TALN), la parole est interprétée. Les intentions du message sont extraites et les variables d'informations (*slots*) identifiées.

6) Un gestionnaire de dialogue permet de préciser le scénario d'interaction à mettre en œuvre avec l'utilisateur en apportant le schéma de réponse approprié.

7) Une réponse adaptée à la requête de l'utilisateur est identifiée et le cas échéant, des ressources distantes sont utilisées : base de connaissance publiquement accessibles (encyclopédie en ligne, etc.) ou par authentification (compte bancaire, application musicale, compte client pour achat en ligne, etc.).

8) Les variables d'informations (*slots*) sont remplies avec les connaissances récupérées.

9) Une phrase de réponse est créée et/ou une action est identifiée (monter les stores, augmenter la température, jouer un morceau de musique, répondre à une question, etc.).

10) Cette phrase est synthétisée (*text to speech*) et/ou l'action à opérer est envoyée à l'équipement.

11) La réponse et/ou la commande est mise en œuvre par l'équipement embarquant l'assistant vocal.

12) L'assistant vocal repasse en veille.

Un assistant vocal n'est donc pas à proprement parler « intelligent ». Les éléments de connaissance proviennent de sources tierces : données en libre accès (encyclopédies en ligne), bases de données contenant des informations renseignées par l'utilisateur (son agenda, son carnet d'adresse, etc.), etc.

> ***« Un assistant vocal n'est donc pas à proprement parler « intelligent ». »***

ZOOM SUR...

Le développement des technologies de traitement automatique de la parole

Suite à l'établissement des bases théoriques du traitement du signal, notamment les théories de l'information et de l'échantillonnage de Claude Shannon, le traitement automatique de la parole (*speech processing*) est devenu une composante fondamentale des sciences de l'ingénieur. Situé au croisement de la physique (acoustique, propagation des ondes), des mathématiques appliquées (modélisation, statistique), de l'informatique (algorithmique, techniques d'apprentissage) et des sciences de l'homme (perception, raisonnement), le traitement de la parole a rapidement été décliné en de nombreux sujets d'étude : identification et vérification du locuteur, transcription automatique de la parole, synthèse vocale, détection des émotions, etc. Depuis une quinzaine d'années, la discipline dans son ensemble a progressé de façon très importante, différents facteurs concourant à cela : amélioration des méthodes, augmentation notable des capacités de calcul et plus grands volumes de données disponibles. L'excellence de la recherche française dans ce domaine est d'ailleurs manifeste avec des laboratoires et centres historiquement reconnus tels qu'entre autres, LIMSI (Paris Saclay), LIUM (Le Mans), LIA (Avignon), LORIA (Nancy), LIG et GIPSA-lab (Grenoble), LPL (Aix-en-Provence), IRIT (Toulouse), Eurecom (Sophia-Antipolis), Ircam, LPP (Paris), etc. (plus d'informations disponibles sur le site de l'AFCP, l'Association francophone de la communication parlée[19]).

La transcription automatique de la parole (*speech to text*) mettait auparavant en œuvre trois étapes distinctes visant à : 1) déterminer quels phonèmes avaient été prononcés à l'aide d'un modèle acoustique ; 2) déterminer quels mots ont été prononcés à l'aide d'un dictionnaire phonétique ; 3) retranscrire la séquence de mots (phrase) ayant le plus de chances d'avoir été prononcée à l'aide d'un modèle de langage. Aujourd'hui, avec les progrès permis par l'apprentissage profond (une technique d'apprentissage automatique), de très nombreux systèmes proposent d'effectuer une transcription automatique de la parole de « bout en bout » (*end to end*). Cela permet d'éviter de passer par l'entraînement complexe de trois modèles différents tout en offrant de meilleures performances tant en termes de résultats que de temps de traitement. Presque tous les grands acteurs du numérique proposent désormais leurs propres implémentations utilisables facilement par des systèmes d'API (*Application Programming Interface)* mais des systèmes *open source* existent également (DeepSpeech[20] ou Kaldi[21] par exemple).

La synthèse de parole met quant à elle en œuvre majoritairement depuis les années 1990, une synthèse dite par concaténation d'unités. Cette technique consiste à sélectionner, dans l'ensemble des enregistrements d'un acteur préalablement transcrits en phonèmes, syllabes et mots, les briques de son qui correspondent aux mots que l'on souhaite faire prononcer par la voix et à les assembler les uns à la suite des autres pour former une phrase intelligible et avec une diction naturelle. L'avantage de cette synthèse est d'être exclusivement basée sur la réutilisation de briques réelles et donc de garantir le naturel de la voix de synthèse. Son désavantage est toutefois d'être limité à la voix de la personne, et à son contenu stylistique et expressif. La synthèse dite statistique ou paramétrique est apparue dès la fin des années 1990 avec les premières tentatives de modéliser les paramètres d'une voix comme l'intonation, le rythme, et le timbre, par des modèles statistiques génératifs comme les chaînes de Markov cachées. Si la synthèse par concaténation est encore largement utilisée, là-encore ce sont désormais les grands acteurs du numérique qui dominent désormais la recherche et développement sur ce secteur avec des réalisations se focalisant sur la synthèse paramétrique comme WaveNet[22], Tacotron[23], DeepVoice[24], ou encore la démonstration de Google Duplex[25] où la voix de synthèse prend un rendez-vous chez le coiffeur.

19 - http://www.afcp-parole.org/
20 - https://github.com/mozilla/DeepSpeech
21 - https://github.com/kaldi-asr/kaldi
22 - Aäron van den Oord et Sander Dieleman, WaveNet: A generative model for raw audio, Deepmind blog, septembre 2016, https://deepmind.com/blog/article/wavenet-generative-model-raw-audio
23 - Yuxuan Wang, Expressive Speech Synthesis with Tacotron, Google AI blog, mars 2018, https://ai.googleblog.com/2018/03/expressive-speech-synthesis-with.html
24 - Deep Voice 3: 2000-Speaker Neural Text-to-Speech, Baidu Research blog, octobre 2017 http://research.baidu.com/Blog/index-view?id=91
25 - Yaniv Leviathan, *Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone*, Google AI blog, mai 2018, https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html

ZOOM SUR...

Le développement des technologies de traitement automatique du langage naturel

Le traitement automatique du langage naturel (TALN) ou *Natural Language Processing* (NLP) en anglais est un domaine multidisciplinaire impliquant la linguistique, l'informatique et l'intelligence artificielle. Il vise à créer des outils de traitement de la langue naturelle pour diverses applications : traduction automatique, génération et résumé automatique de textes, correction orthographique, systèmes de questions-réponses, fouille de textes, reconnaissance d'entités nommées, analyse de sentiments, etc. La notion prend son essor dès les années 1950 avec les travaux d'Alan Turing et son fameux test visant à caractériser si dans un échange de messages écrits, un sujet humain peut déterminer s'il s'adresse à une machine ou non[26].

Concrètement, l'objectif du TALN est de donner aux machines la capacité de donner un sens aux échanges humains. Il s'agit là d'un défi important car les outils informatiques exigent traditionnellement qu'on interagisse avec eux dans un langage de programmation formel, c'est-à-dire précis, non ambigu et très structuré. Cependant, la parole humaine peut être imprécise, équivoque et sa structure peut varier en fonction du niveau de langage utilisé, du domaine d'application, etc. Les deux principales branches du TALN sont l'analyse syntaxique et l'analyse sémantique. La première permet d'évaluer le sens d'une phrase en fonction de règles grammaticales. Il s'agit ainsi de s'appuyer sur l'arrangement des mots qui la constituent. Parmi les techniques utilisées, citons la segmentation des mots (qui divise un texte en unités), la rupture de phrase (qui place les limites de la phrase d'un texte), la segmentation morphologique (qui divise les mots en fonction de leur composition) et la racinisation (qui regroupe ensemble les mots ayant une racine commune). L'analyse sémantique utilise la signification des éléments constitutifs d'un texte. La désambiguïsation du sens des mots (qui dérive la signification d'un mot en fonction du contexte) et la reconnaissance des entités nommées (qui consiste à rechercher des objets textuels comme des noms propres, dates, lieux, etc.) sont des techniques fréquemment utilisées.

Alors que les premières approches de TALN étaient fondées sur l'utilisation d'algorithmes d'intelligence artificielle se reposant sur des systèmes de règles et ontologies, les approches actuelles emploient des méthodes d'apprentissage automatique et plus spécifiquement l'apprentissage profond (*deep learning*). Là encore, la conjonction de l'amélioration des méthodes développées, de l'augmentation des données disponibles et du développement des capacités calculatoires ont permis la mise en œuvre d'approches basées sur l'apprentissage automatique pour effectuer des inférences statistiques à partir de l'analyse de très larges corpus textuels. Bien souvent les laboratoires et centres de recherche français travaillant sur les thématiques du traitement automatique du langage naturel sont également impliqués dans le traitement automatique de la parole. On retrouve donc les mêmes acteurs que ceux indiqués dans l'encadré précédent. Peuvent toutefois également être cités, sans que la liste soit exhaustive : LS2N (Nantes), LIS (Marseille), IRISA (Rennes), GREYC (Caen), LIRMM (Montpellier), etc. (plus d'informations disponibles sur le site d'ATALA l'association pour le traitement automatique des langues[27]).

Un assistant vocal, c'est qui ?

L'assistant vocal implique cinq catégories d'acteurs (voir infographie page 14) :

• **Les concepteurs :** responsables du développement de l'assistant vocal, ils conçoivent et définissent son fonctionnement et ses possibilités : modalités d'activation, choix d'architecture, accès aux données, gestions des enregistrements, spécifications matérielles, etc.

• **Les développeurs d'applications :** à la manière de ce qui se fait pour les applications mobiles, ils souhaitent développer une application tierce à destination d'un assistant. Pour cela, il est nécessaire de respecter les contraintes de développement imposées par le concepteur.

• **Les intégrateurs :** constructeurs d'objets et équipements

26 - Alan Turing, *Computing machinery and intelligence*, Oxford University Press, vol. 59, no 236, 1950
27 - https://www.atala.org/

ZOOM SUR...

Les données, levier essentiel pour le développement d'un assistant vocal

De multiples briques technologiques sont nécessaires au bon fonctionnement d'un assistant vocal. Historiquement, la recherche scientifique française et internationale s'est appuyée sur la production de nombreux corpus afin afin de développer et mesurer les avancées en matière de reconnaissance du locuteur, de transcription automatique, de détection de mots-clés, etc. L'Institut des standards et technologies aux États-Unis (NIST) et la Direction générale de l'armement (DGA) en France ont été à l'initiative de vastes campagnes d'évaluation avec les programmes Speech Analytics[28] et Speaker and Language Recognition[29] pour le premier et les campagnes ESTER, ETAPE ou encore REPERE pour la seconde. Enfin, depuis les années 1990, plusieurs organisations telles que la European Language Resources Association[30] ou le Linguistic Data Consortium[31] proposent de rassembler et distribuer des ressources linguistiques orales, écrites et terminologiques afin de favoriser le développement des technologies de traitement automatique de la voix et de la parole.

D'importants progrès ont été enregistrés rendant finalement possible, à partir des années 2010, le développement d'assistants vocaux à destination du grand public. Toutefois, à quelques exceptions près (comme par exemple le modèle de langage BERT[32] mis à disposition par Google et dont il existe une déclinaison française CamemBERT[33]), la plupart des données utilisées par les grandes entreprises du numérique en position prédominante ne sont pas aisément disponibles. Plusieurs initiatives peuvent toutefois être pointées telles que le projet Common Voice[34] initié par Mozilla, visant à recueillir des enregistrements audio dans de nombreuses langues, ou encore le Voice Lab[35], association française regroupant une trentaine d'acteurs dont l'objectif est également de constituer des ressources vocales prenant en compte les accents ainsi que les dialectes régionaux et internationaux. La diversité linguistique, notamment à des fins d'inclusion, doit en effet être prise en compte alors que les grands acteurs du marché se concentrent généralement sur les langues, dialectes et accents considérés comme les plus rentables. La puissance publique elle-même se saisit d'ailleurs du sujet de la constitution de ressources langagières avec le projet PIAF[36].

connectés, ils souhaitent doter ces derniers d'un assistant vocal. Ils s'assurent pour cela que les spécifications minimales définies par les concepteurs sont bien respectées.

• **Les déployeurs :** responsables de lieux accueillants des personnes (lieux d'hébergement, environnements professionnels, véhicules de location, etc.) ils souhaitent mettre à disposition de leur public des assistants vocaux (éventuellement avec des applications dédiées).

• **Les utilisateurs :** maillon final de la chaîne de valeur des assistants vocaux, ils peuvent employer ces derniers sur différents équipements (enceinte, téléviseur, smartphone, montre, etc.).

En fonction des modèles d'affaires (voir les stratégies des acteurs page 24) et des choix technologiques, certains acteurs peuvent endosser plusieurs combinaisons de rôles, par exemple être concepteur et intégrateur, concepteur et développeur d'application, etc.

Un assistant, pour quoi faire ?

On compte plusieurs avantages à passer par la parole comme : le caractère naturel de l'interaction qui n'implique pas d'apprentissage spécifique, la rapidité d'exécution de la commande qui avoisine les performances d'un expert pour la frappe au clavier et l'extension du champ d'action qui peut permettre d'avoir accès plus rapidement à une information. Si le fait de se reposer sur la parole emporte

28 - https://www.nist.gov/programs-projects/speech-analytics
29 - https://www.nist.gov/programs-projects/speaker-and-language-recognition
30 - http://www.elra.info/en/
31 - https://www.ldc.upenn.edu/
32 - Jacob Devlin et Ming-Wei Chang, *Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing*, Google AI blog, novembre 2018, https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html
33 - https://camembert-model.fr/
34 - https://voice.mozilla.org/fr
35 - http://www.levoicelab.org/
36 - https://piaf.etalab.studio/

également des difficultés pour que soit interprété correctement le message (variabilités du signal audio entre les différents locuteurs, de l'environnement acoustique, ambiguïté de la langue, etc.), ses bénéfices ont également été clairement identifiés.

En pratique, la fluidification ou la simplification de tâches reste la première des motivations pour s'équiper en assistants vocaux. Il peut s'agir par exemple de passer/répondre à un appel, lancer un minuteur, etc., en particulier lorsque l'utilisateur a ses mains indisponibles, parce qu'il fait la vaisselle, s'habille ou bricole. La maison connectée et la domotique sont les grandes applications mises en avant par les concepteurs d'assistants vocaux. En proposant de simplifier l'exécution de tâches (allumer la lumière, régler le chauffage, baisser les volets, etc.) et de les centraliser à travers un seul et unique outil facilement activable à distance et sans intermédiaire, ils s'inscrivent dans les discours comme un facilitateur domestique. Outre les usages personnels ou dans le cadre du foyer, l'usage de la commande vocale peut présenter un intérêt dans les environnements professionnels dans lesquels il est difficile de pouvoir manipuler des outils informatiques et utiliser des commandes écrites (travail ouvrier de manufacture par exemple).

De grands bénéficiaires de l'interface vocale pourraient en théorie être les publics en situation de handicap ou de dépendance pour lesquels l'usage des interfaces traditionnelles pose problème. Autrement dit, l'assistance vocale peut permettre un accès facilité à l'information et aux ressources informatiques et ainsi promouvoir des logiques d'inclusion. La sociologue Dominique Pasquier souligne notamment que le passage par la voix permet de dépasser les difficultés liées à l'écrit, que l'on peut retrouver chez les classes populaires[37].

Enfin, la santé est également un domaine porteur de nombreux cas d'usages pour les agents conversationnels, qu'ils soient vocaux ou non comme l'indique dans son état des lieux le Lab e-santé, think tank spécialisé dans les sujets de santé numérique[38]. Si les cas d'usages sont encore peu nombreux en France, pour certains, c'est tout le parcours de soin du patient qui pourrait, à terme, être impacté par les interactions homme/assistant : non seulement pour le bien-être et la prévention, mais aussi pour le curatif et l'accompagnement dans le traitement. On observe d'ailleurs de nombreux partenariats noués entre grands acteurs du numérique et professionnels du monde médical (voir encadré page 26).

ZOOM SUR...

Les séniors, premiers bénéficiaires de ces nouveaux outils ?

Les dernières études démographiques indiquent que si en 2020, environ 19,6 % de la population française est constituée de personnes d'au moins 65 ans, cette proportion va fortement augmenter. Selon les estimations de l'INSEE elle devrait atteindre environ un quart de la population en 2040[39]. Cette évolution sociétale introduit un nombre important de défis liés aux besoins spécifiques des personnes âgées, dont l'avancée dans l'âge s'accompagne souvent d'une fragilisation de l'état de santé, de difficultés à poursuivre une vie sociale, voire, dans certains cas, à accomplir des actes de la vie quotidienne.

Les assistants vocaux, dont l'utilisation ne nécessite précisément pas de manipuler un clavier, un écran ou une souris, et dont l'utilisation ne requiert, du moins en principe, que l'utilisation de commandes vocales proches du langage naturel, peuvent précisément offrir une promesse de simplicité et palier certains handicaps, ce qui en a fait un objet privilégié de recherche en matière de « silver économie ». Plusieurs offres sont actuellement commercialisées dans l'hexagone. On peut citer à titre d'exemple le smartphone OLGA (Flagtory SAS) et les enceintes SkipIt (consortium HomeKeeper) et Fanny (Dynséo). Ces dispositifs, dont les architectures techniques varient, sont dotés d'une interface comprenant un assistant vocal. Celle-ci permet de gérer des fonctionnalités de base sans contact physique avec l'appareil : écrire et écouter un message, émettre un appel, mettre la radio et jouer de la musique. Enfin, de nombreux modules spécialisés pouvant être intégrés aux assistants vocaux « classiques » (Siri, Alexa, Google Assistant, etc.) existent et ciblent les besoins des personnes âgées à travers des applications de jeux de mémoire, de lecture avec une voix de synthèse, ou encore l'appel de numéros d'urgence.

37 - Dominique Pasquier, *Dans les classes populaires, la vie privée relève moins de l'individu que du groupe familial*, Linc.cnil.fr, mars 2020, https://linc.cnil.fr/dominique-pasquier-dans-les-classes-populaires-la-vie-privee-releve-moins-de-lindividu-que-du-groupe
38 - Lab e-santé, Chatbot : *le futur de la santé passe-t-il par le conversationnel ?*, juillet 2019, https://www.ticsante.com/documents/201907041716270.Livre_blanc_chatbot_du_Lab-esante.pdf
39 - INSEE, *Tableaux de l'économie française* (édition 2018), fiche 3.2 « Population par âge », https://www.insee.fr/fr/statistiques/3303333?sommaire=3353488

QUELS USAGES DES ASSISTANTS VOCAUX ?

Pour les concepteurs d'assistants vocaux, la promesse de ces dispositifs est bien souvent de jouer un rôle de majordome du foyer connecté. Toutefois, les usages actuels semblent encore balbutiants et il est légitime de se demander en citant l'enseignant-chercheur en design Anthony Masure : « Si les assistants vocaux sont la solution, quel est le problème ? »[40].

Des dispositifs de plus en plus nombreux...

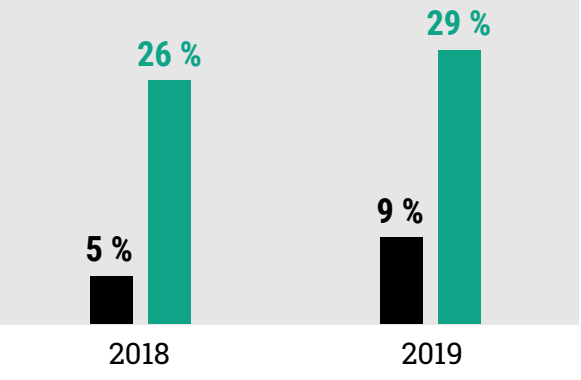
Depuis plusieurs mois, de nombreux chiffres et projections tablent sur une diffusion grandissante de ces nouveaux outils numériques. On évoque même une tendance comparable à celle qu'ont connue les tablettes. À l'été 2019, il était fait état d'environ 1,7 million d'assistants vocaux sur enceintes connectées et d'un usage de l'assistant vocal du smartphone par 16 à 20 millions d'utilisateurs en France. Dans le monde, le cabinet de conseil Roland Berger estime à 3 milliards le nombre d'assistants vocaux, tout support confondu, et présente une projection de 8 milliards en 2023[41]. Ces chiffres pourraient encore augmenter grâce à l'intégration toujours accrue de ces assistants dans des objets du quotidien, à commencer par la voiture connectée.

Depuis 2018, la CNIL inclut dans le sondage annuel réalisé avec Médiamétrie et portant sur les pratiques numériques des Français, une question relative aux assistants vocaux[42, 43]. À la lueur des réponses fournies par les quelques 2 000 personnes interviewées, plusieurs enseignements peuvent être tirés. Tout d'abord, les usages restent encore discrets avec un tiers des personnes interrogées qui déclarent avoir utilisé un assistant vocal dans les 12 derniers mois. Ces indicateurs sont néanmoins en augmentation pour 2019. L'utilisation de l'assistant du smartphone passe de 26 à 29 % quand l'utilisation d'un assistant vocal sur enceinte connectée est proche de doubler, passant de 5 à 9 % (voir figure 1). En parallèle, 20 % des internautes interrogés déclarent avoir désactivé l'assistant personnel présent sur leur smartphone. Enfin, les usages se précisent et on peut noter un paramétrage plus grand de ces objets. 46 % des personnes interrogées et ayant utilisé ce genre d'outils (soit environ 700 personnes) indiquent ainsi avoir déjà procédé à un paramétrage de leur assistant vocal : vérifier la configuration, supprimer l'historique des commandes vocales passées, etc. On observe ainsi une augmentation de 6 points par rapport à ce qui avait été mesuré l'année précédente.

Figure 1

Évolution de l'utilisation des assistants vocaux en France entre 2018 et 2019

(Base : internautes de 15 ans et plus équipés de smartphones, ordinateur ou enceinte connectée (n=2 102))



40 - Hubert Guillaud, *« Si les assistants vocaux sont la solution, quel est le problème ? »*, InternetActu blog LeMonde.fr, janvier 2019, https://www.lemonde.fr/blog/internetactu/2019/01/13/si-les-assistants-vocaux-sont-la-solution-quel-est-le-probleme/

41 - Cabinet Roland Berger, La révolution naissante des assistants vocaux, juillet 2019, https://www.rolandberger.com/fr/Publications/La-r%C3%A9volution-naissante-des-assistants-vocaux.html

42 - LINC, *[Baromètre LINC 2019] - Les pratiques de protection des données progressent*, Linc.cnil.fr, décembre 2019 https://linc.cnil.fr/fr/barometre-linc-2019-les-pratiques-de-protection-des-donnees-progressent

43 - LINC, *[Baromètre LINC 2018] - Des utilisateurs plus passifs vis-à-vis des assistants vocaux que des smartphones ou navigateurs*, Linc.cnil.fr, octobre 2018 https://linc.cnil.fr/fr/barometre-linc-des-utilisateurs-plus-passifs-vis-vis-des-assistants-vocaux-que-des-smartphones-ou

Le rapport *Assistants vocaux et enceintes connectées* du CSA et de la Hadopi de mai 2019 revient également sur les volontés d'équipement pour les personnes ne disposant pas d'enceintes connectées[44]. Selon leurs estimations, seulement 4 % d'entre eux seraient prêts à en acheter une. Les raisons sont d'abord « l'inutilité » de ces dispositifs (67 %), mais également la sécurité des données personnelles (59 %). Dans un de ses rapports, la société Microsoft tend à confirmer cet enjeu[45]. Celui-ci pointe aussi les questions relatives à la sécurité des dispositifs, à la protection des données personnelles et aux possibilités d'espionnage via l'écoute passive. Au contraire, les personnes déjà équipées voient les assistants vocaux comme une « vraie innovation » qui va « révolutionner le quotidien » (pour 53 % d'entre eux). Enfin, le baromètre annuel de l'ARCEP montre que les plus jeunes sont les plus à même d'apprécier ce genre d'équipements, et que ces assistants sont plus répandus dans les foyers nombreux[46].

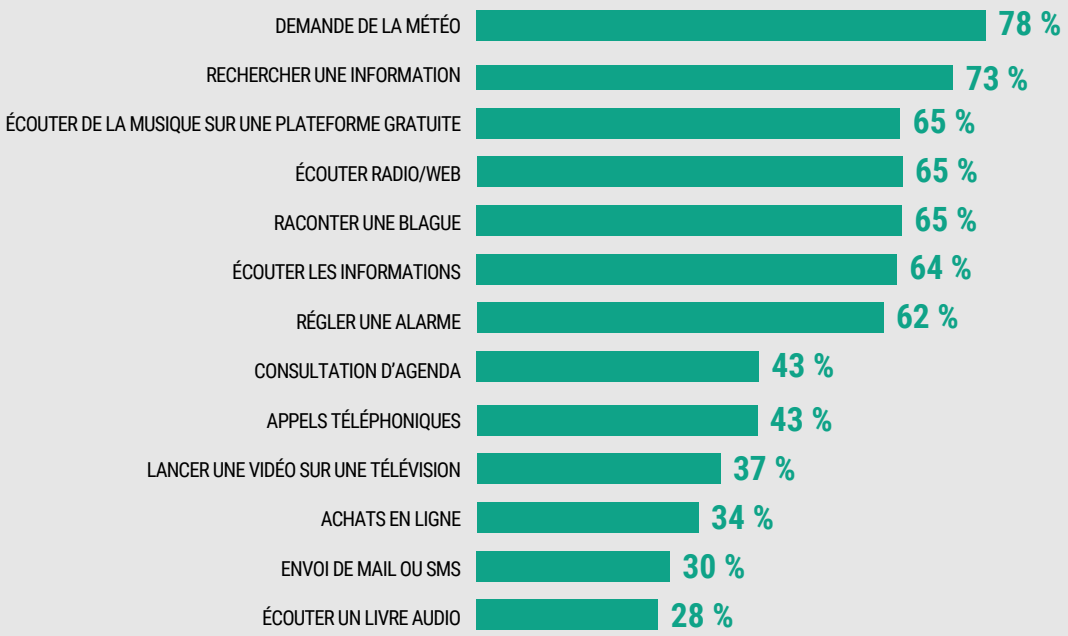
… pour des usages encore prudents

D'après le Baromètre LINC 2019, seuls 33 % des internautes ont utilisé un assistant vocal sur les 12 derniers mois. Les usages sont concentrés sur la « facilitation de petits gestes ». Sur ce dernier point, le rapport publié par le cabinet d'audit anglophone PwC précise que « la majorité des objets achetés sont petits et légers, généralement quelque chose que quelqu'un pourrait acheter sans avoir besoin de le voir (pour en déterminer la qualité par exemple) »[47]. Concrètement, lorsque l'assistant est utilisé pour effectuer une dépense, il s'agit de petits achats, de commandes de repas ou de petites courses. L'enquête présentée dans le rapport CSA–Hadopi conforte cette idée de légèreté dans les usages (voir figure 2), et d'un recours plus rare aux usages plus intrusifs (achats, dictée de sms ou mails, etc.). Impression également confirmée en regardant quelles sont les applications les plus téléchargées par les utilisateurs. Dans le cas d'Amazon Alexa, il s'agit avant tout de contenus radio, de jeux ou d'environnements sonores (sons d'animaux ou sons de l'océan).

Figure 2

Quels usages pour les enceintes équipées d'assistants vocaux

(Source rapport CSA–Hadopi, utilisateurs d'enceintes connectées au cours des 30 derniers jours au moment de l'enquête, soit 287 individus).



44 - CSA-Hadopi, *Assistants vocaux et enceintes connectées - L'impact de la voix sur l'offre et les usages culturels et médias*, mai 2019, https://www.csa.fr/Informer/Collections-du-CSA/Thema-Toutes-les-etudes-realisees-ou-co-realisees-par-le-CSA-sur-des-themes-specifiques/Les-autres-etudes/Etude-HADOPI-CSA-Assistants-vocaux-et-enceintes-connectees

45 - *Microsoft, Voice report : From answers to action: customer adoption of voice technology and digital assistants*, 2019 https://advertiseonbing-blob.azureedge.net/blob/bingads/media/insight/whitepapers/2019/04%20apr/voice-report/bingads_2019voicereport.pdf

46 - Arcep, *Le baromètre du numérique*, novembre 2019, https://www.arcep.fr/cartes-et-donnees/nos-publications-chiffrees/numerique/le-barometre-du-numerique.html

47 - PwC, Report : *Prepare for the voice revolution*, février 2019, https://www.pwc.com/cisvoiceassistants

Quels usages pour demain ?

[Design Fiction]

CNIL — **2021** — Mise à disposition sur les grands magasins d'applications (*stores*) d'une application *Besoin d'aide* à destination du grand public

RATP — **2022** — Déploiement d'assistants vocaux dans les arrêts de bus RATP afin consulter l'état du trafic en temps réel et se renseigner sur son itinéraire

CPAM — **2023** — Lancement de l'opération « Mon médecin traitant est un *chatbot* » qui propose une consultation préalable avec un assistant vocal avant toute prise de rendez-vous médical

Pôle Emploi — **2024** — Mise en œuvre d'assistants vocaux dans les centres Pôle Emploi afin de faciliter la primo inscription des nouveaux demandeurs

Medical Robotics Japan* — **2027** — Première étude clinique prouvant l'effet positif d'un assistant vocal dédié sur la survie de personnes âgées maintenues à domicile

Startup Dial'Og[48]* — **2027** — Levée de 100 millions de dollars pour sa sonde de l'opinion, *« l'arme absolue des élus qui écoutent à l'oreille des électeurs »*. Un système de reconnaissance vocale pour révolutionner la participation citoyenne.

MIT-Inria — **2029** — Design d'un mot d'activation non-ambigu et utilisable avec tous les assistants permettant ainsi d'éviter les fausses activations

Commission européenne — **2030** — Adoption d'un règlement encadrant strictement les commandes pouvant être réalisées par un assistant vocal dans un véhicule suite à une série d'accidents mortels provoqués par une incompréhension avec l'utilisateur

Mind Technologies* — **2046** — Premier modèle industriel pour un « assistant de pensée » pouvant être commandé avec son esprit.

* Ce nom a été inventé

48 - LINC, Cahier IP7, *Civic tech, données & Demos,* « Quand notre voix est entendue », p. 30, décembre 2019. https://linc.cnil.fr/fr/civic-tech-donnees-et-demos-le-cahier-ip7-explore-les-liens-entre-democratie-et-technologies

QUELLE(S) STRATÉGIE(S) POUR LES CONCEPTEURS D'ASSISTANTS VOCAUX ?

Dès 2018, le LINC entrevoyait que le positionnement des acteurs sur le marché des assistants vocaux s'inscrivait d'abord dans le développement de leurs modèles économiques préexistants (publicité, sites marchands, magasins d'applications, etc.), avec un marché structuré autour des géants du numérique américains, coréens ou chinois[49]. Ces grands acteurs reproduisent ou étendent la stratégie de plateforme qui a fait leur succès : ils multiplient les services gratuits (à destination des individus comme des entreprises) pour capter le plus d'utilisateurs possibles par effets de réseaux. En périphérie, gravitent des acteurs plus spécialisés qui tentent de se distinguer par leurs performances ou des visées plus précises en termes d'intégration de leurs produits.

Des stratégies diverses en fonction des acteurs

En fonction de leurs métiers historiques, les professionnels concevant des assistants vocaux présentent des positionnements économiques différents.

• Une fonctionnalité levier pour la vente de smartphones ou de logiciels

Pour les constructeurs de smartphones ou développeurs de systèmes d'exploitation – qu'il s'agisse d'Apple, de Samsung, de Huawei ou de Google – le fait d'avoir un assistant vocal intégré constitue d'abord un avantage concurrentiel sur la marché de la vente de smartphones. Le lancement de Siri a contribué au succès des iPhone et iPad. Depuis, les assistants vocaux se multiplient : Google Assistant, Samsung Bixby, Microsoft Cortana... L'intérêt est d'abord de déployer, dans une sorte d'extension du kit main libre, l'usage du smartphone. Cette fonctionnalité a déjà dépassé la nouveauté technologique et constitue désormais l'offre de base et n'est plus une surprise du côté des utilisateurs. La présence d'un assistant devient donc un avantage concurrentiel. Samsung a sorti les versions anglaise et française de Bixby en 2018 et Huawei a annoncé que son assistant Xiaoyi serait déployé à l'international. Le constructeur chinois a même dévoilé un nom pour la version française : Célia. On observe donc une logique forte de pénétration de marché et de ventes de produits à l'international.

• Des pourvoyeurs de clients sur les places de marché

Les grandes plateformes que sont Google et Amazon transforment leurs assistants vocaux en bases avancées de leurs comparateurs de prix et *marketplaces*. Il s'agit d'un nouvel intermédiaire pour leurs activités de vente en ligne, qu'il s'agisse respectivement de Google Shopping ou du site d'Amazon. Ainsi, lorsqu'une commande de courses est passée via l'assistant, elle se répercutera d'abord dans leur écosystème de vente en ligne. L'assistant proposera notamment les offres référencées sur son propre site. L'enjeu étant pour ces acteurs d'intégrer au plus près l'acte d'achat dans le quotidien des individus, et cela de la manière la plus fluide possible. Cet enjeu de fluidité est par la suite partagé avec les développeurs d'applications marchandes tierces. On parle alors de v-commerce, pour ce nouvel avatar de l'e-commerce disponible uniquement par l'interface vocale. Être positionné en intermédiaire, mettant en relation un individu avec des fournisseurs de contenus, est le métier historique de grandes plateformes telles que Google et Amazon. Le rapport CSA-Hadopi revient notamment sur cette place prise dans le secteur des industries culturelles, et la nouvelle relation des éditeurs de contenus et des concepteurs d'assistants vocaux : « [Google, Amazon et Apple] ont, dès leur lancement, cherché à proposer une offre attractive de médias et de contenus culturels, en nouant des partenariats avec des éditeurs ou en mettant en avant leur propres services ».

49 - Olivier Desbiey, Ok Google et Siri ne suivent pas la même voie qu'Alexa ou Cortana, Linc.cnil.fr, mars 2018, https://linc.cnil.fr/fr/ok-google-et-siri-ne-suivent-pas-la-meme-voie-qualexa-ou-cortana

• Des sources de données pour le profilage et la publicité ciblée

L'utilisation des assistants intelligents requiert dans de très nombreux cas un moment de paramétrage dans lequel il est indispensable de lier un compte à l'assistant. Les concepteurs peuvent ainsi enrichir le profil de leurs utilisateurs à travers l'utilisation de l'assistant, les applications (ou compétences) qui sont installées, des commandes passées, etc. On retrouve ici l'application classique du modèle de marché biface : gratuit du côté utilisateur, mais payé par les annonceurs afin d'affiner leur capacités de prospection commerciale[50]. Ces derniers pourront alors cibler des personnes qui correspondent à un certain type de profil qu'ils jugeront comme étant le plus à même d'acheter leurs produits. Par exemple, dans le cas de Google, les interactions avec l'assistant sont une nouvelle source d'informations rattachée à un compte utilisateur qui regroupe également des informations relatives aux recherches effectuées sur le moteur de recherche, aux actions sur téléphone Android, vidéos visionnées sur YouTube, ou itinéraires de navigation sur Maps.

• Des stratégies divergentes : Microsoft et Apple

Même parmi les grands acteurs du numériques, différents positionnement peuvent être observés. Ainsi, l'assistant Cortana est l'exemple de la logique suivie par Microsoft : se tourner vers le monde professionnel plutôt que les particuliers. L'entreprise a d'ailleurs annoncé qu'il n'y aura plus d'application de Cortana sur Android et iOS à partir du 31 janvier 2020[51]. Pour autant, l'assistant est loin d'être abandonné. Cortana sera intégré dans les services d'Office 365 de Microsoft, ainsi que dans les produits du groupe, à l'instar d'Outlook. La stratégie de Microsoft est d'équiper ses produits, faisant ainsi de Cortana un outil à part entière des différents supports Windows. De son côté, si Apple a été pionnier dans le développement et l'installation de son assistant vocal Siri sur l'ensemble de ses produits, son enceinte connectée – le HomePod – n'est arrivée que tardivement sur le marché en comparaison des autres grands fournisseurs cités plus haut (2018 en France et aux États-Unis).

Une logique économique encore différente est ici à l'œuvre avec un produit tourné vers la musique et la qualité de son et s'inscrivant dans un écosystème fermé, à l'image des autres produits Apple. Le HomePod ne fonctionne ainsi pour l'instant qu'avec les applications Apple. La firme de Tim Cook suit donc une autre stratégie que celle de la diffusion massive et de la facilité d'accès de ses concurrents.

• Des logiques de fourniture de solutions en marque blanche

Enfin, certains acteurs choisissent de se positionner sur des modèles B2B (*business to business*). Il s'agit de proposer via des logiques de « marque blanche » de fournir un assistant vocal « sur-mesure » à intégrer dans l'équipement du client. Cet assistant est ensuite commercialisé sous la marque de l'acheteur. Ce modèle économique est souvent celui de (plus) petits acteurs. Il s'agissait notamment de celui de la startup française Snips, rachetée par Sonos à la fin de l'année 2019, ou encore de celui de la société Nuance Communications (entreprise qui a travaillé en partenariat avec Apple sur les premières versions de Siri) via son assistant Nina. C'est aussi le cas de Harman, filiale de Samsung, qui équipe des paquebots de croisière en systèmes sonores et lumineux et a développé Zoe, un assistant dédié à la croisière, installé dans toutes les cabines de certains paquebots de la société MSC Croisières[52].

La nouvelle frontière de la stratégie de plateforme des acteurs du net

Les grands constructeurs d'assistants vocaux comme Google ou Amazon ont fait le choix d'une stratégie de diffusion à grande échelle. Pour cela, ils mettent en avant les dizaines ou centaines de millions (voire milliards !) d'équipements adressables par leur assistant. Outre cette force de frappe commerciale déjà présente, le développement s'est opéré par deux moyens : 1) le déploiement très facile de l'assistant sur tout support équipé d'un microphone, d'un haut-parleur, d'une interface réseau et de quelques ressources calculatoires, et 2) la facilité de développer une application tierce, sur un modèle de magasin d'applications mobiles. La fourniture de la documentation et d'un kit de développement logiciel (SDK pour *Software Development Kit*), permet à tout développeur de s'intégrer facilement à l'écosystème de l'assistant. Toutefois, si les applications peuvent être construites par tout le monde, elles ne peuvent l'être qu'avec les outils limités et normés mis à disposition par le constructeur. Jusqu'à présent, ce modèle de fonctionnement est gratuit. La question de possibles évolutions, par exemple basées sur des modèles marchands plus classiques, se pose toutefois. Ainsi, un système de paiement pour le référencement des applications est-il à prévoir dans le futur ?

50 - Jean-Charles Rochet et Jean Tirole, *Plateform Competition in two-sided markets*, Journal of the European Economic Association, juin 2003, https://www.tse-fr.eu/articles/platform-competition-two-sided-markets
51 - Microsoft, *Changes to Cortana services*, novembre 2019, https://support.microsoft.com/en-gb/help/4531683
52 - Harman, Real-time AI: Meet ZOE, the cruise industry's world's first voice-activated digital assistant, janvier 2019, https://services.harman.com/Blogs/real-time-ai--meet-zoe-the-cruise-industry%E2%80%99s-world%E2%80%99s-first-voice-activated-digital-assistant

ZOOM SUR...

Une diversification des usages à travers des partenariats

Conséquence de stratégies de dissémination sectorielles, les assistants vocaux se retrouvent dans des domaines pouvant toucher tout autant à la vie quotidienne qu'à l'intimité, comme par exemple celui de la santé. Dans ce domaine-là, des partenariats entre concepteurs d'assistants et acteurs du secteur – publics autant que privés – ont vu le jour. On peut notamment citer celui signé entre Amazon et la NHS (National Health Service), le service de santé du Royaume-Uni. Ce dernier a soulevé des critiques de la part d'acteurs de la société civile, réclamant plus de transparence dans le fonctionnement de ce partenariat. L'association Privacy International s'est ainsi interrogée sur l'association de données de santé émanant d'un service public et le modèle économique d'Amazon qui repose sur la récupération de données à des fins publicitaires[53]. L'assistant d'Amazon a également été le support du partenariat Alexa Diabetes Challenge, passé entre la firme de Jeff Bezos et le laboratoire pharmaceutique Merck[54]. En l'occurrence, il s'agissait d'encourager de jeunes startups à proposer des solutions d'aide ou d'accompagnement des personnes diabétiques se reposant sur l'assistant vocal Alexa.

Dans le cas des contenus des acteurs culturels, le rapport CSA-Hadopi a souligné le choix de la part de Google de multiplier les partenariats avec les médias. Des accords passés « notamment avec les éditeurs de contenus audio, pour pouvoir proposer une offre d'information (partenariats avec Europe 1, Radio France, RTL, L'Équipe, BFM Business [...]) », visant à promouvoir les produits et services de l'entreprise de Mountain View, qui précise toutefois ne rechercher ni exclusivité, ni exclusion d'autres services. Dans un cadre professionnel, Microsoft noue des partenariats, comme avec le chinois Xiaomi[55]. De façon similaire, Apple et Salesforce ont fait cause commune afin notamment d'implémenter les outils de l'assistant comme Siri Shortcuts dans les produits développés par l'éditeur de logiciel de gestion de la relation client (*Customer Relationship Management*, CRM)[56].

Le partenariat avec des grandes entreprises ou institutions constitue donc un autre moyen de dissémination, permettant d'intégrer des assistants à des produits déjà existants et/ou populaires. En France, une illustration parlante est celle de l'opérateur et fournisseur d'accès Free, qui fin 2018 a intégré l'assistant vocal Alexa directement dans sa box Internet Freebox Delta. La box contient désormais un microphone, un haut-parleur, quelques capacités calculatoires pour la détection du mot-clé, et également un bouton physique permettant de couper le microphone. Toutefois, la multiplication des partenariats pourrait soulever des questions de neutralité des terminaux. Dans un dossier publié en juin 2018, l'Autorité de régulation des communications électroniques, des postes et de la distribution de la presse (ARCEP) fait l'analyse du rôle de ces nouveaux intermédiaires : « les smartphones, les assistants vocaux, les voitures connectées et autres terminaux se révèlent être le maillon faible de l'ouverture Internet. Parce qu'ils ne sont donc pas neutres et peuvent limiter la liberté des utilisateurs de choisir les contenus et services sur Internet »[57]. Dès lors, ce nouveau point d'entrée, à travers son fonctionnement en écosystème fermé et propriétaire, contraint l'utilisateur à accéder aux services selon des règles fixées par le fournisseur de l'assistant.

53 - Privacy International, *Alexa, what is hidden behind your contract with the NHS?*, décembre 2019, https://privacyinternational.org/node/3298
54 - Alexa Diabetes Challenge, http://www.alexadiabeteschallenge.com/
55 - LeBrief, *Xiaomi et Microsoft s'associent, Cortana pourrait être utilisé dans une enceinte connectée du chinois*, Next Inpact, février 2018, https://www.nextinpact.com/brief/xiaomi-et-microsoft-s-associent--cortana-pourrait-etre-utilise-dans-une-enceinte-connectee-du-chinois-2824.htm
56 - Apple Newsroom, *Apple et Salesforce associent les meilleurs appareils pour les entreprises à la solution CRM n° 1 dans le monde*, septembre 2018, https://www.apple.com/fr/newsroom/2018/09/apple-and-salesforce-partner-to-help-redefine-customer-experiences-on-ios/
57 - Arcep, *L'influence des terminaux sur l'ouverture d'internet*, juin 2019 https://www.arcep.fr/la-regulation/grands-dossiers-internet-et-numerique/linfluence-des-terminaux-sur-louverture-dinternet.html

« Les usages en lien avec la domotique de la maison sont pour l'heure encore peu développés mais jugés très intéressants et porteurs d'une véritable valeur ajoutée pour le quotidien des utilisateurs. »

(SOURCE RAPPORT CSA-HADOPI, ASSISTANTS VOCAUX ET ENCEINTES CONNECTÉES, MAI 2019)

Au-delà des ordiphones et des enceintes dédiées, les assistants vocaux ont donc vocation à être intégrés de manière de plus en plus poussée dans les objets domestiques, afin de répondre à des logiques plus concrètes et utilitaires et de permettre à leurs concepteurs d'entrer sur de nouveaux marchés. De nombreux exemples existent déjà : réfrigérateur connecté Samsung Family Hub embarquant l'assistant vocal Bixby[58], thermostat Nest Guard (filiale de Alphabet/Google)[59], robot ménager Monsieur Cuisine connect de Lidl[60], etc. Les deux derniers exemples sont emblématiques de la volonté de pollinisation des concepteurs d'assistants. Ils ont tous deux causé scandale puisqu'il s'est avéré qu'ils intégraient tous deux un microphone, alors qu'il n'en était pas fait mention dans la documentation qui accompagnait les produits. En effet, ce dernier n'était pas « actif », car les services proposés par les deux dispositifs n'intégraient pas de commande vocale… au moment de leur vente. Toutefois, dans une perspective de mise à jour et d'intégration à un service plus large, les constructeurs se tenaient prêt à déployer un assistant vocal au moment opportun.

ZOOM SUR…

L'espion qui me logeait

En décembre 2017, la journaliste Kashmir Hill a choisi d'équiper complètement sa maison d'objets connectés. Elle a tiré un article de cette expérience intitulé *The House that Spied on Me*[61]. L'équipe du LINC avait alors résumé le but de l'expérimentation et ses conclusions. L'objectif était de documenter et comprendre ce qu'il se passe lorsque l'on va au bout du tout connecté, et en particulier en ce qui concerne la protection des données.

« Dans cette maison intelligente, chacun des objets connectés agissait à la manière d'un traceur installé sur notre navigateur, en mesure, si on sait l'utiliser, de déduire des habitudes du foyer, et ses modes de consommation. Ces informations pourraient grandement intéresser des *data brokers*, des agences marketing voire d'autres types d'acteurs, raison pour laquelle il est nécessaire d'accompagner les utilisateurs dans la compréhension du fonctionnement de ces objets, de veiller à la mise en place de moyens d'information et de collecte du consentement « libre, éclairé et révocable » adaptés à ces interfaces ».

> Article à retrouver sur LINC[62]

58 - Samsung, *Family Hub*, https://www.samsung.com/fr/familyhub/
59 - C. Scott Brown, *Nest Secure has an unlisted, disabled microphone*, Android Authority, février 2019, https://www.androidauthority.com/nest-secure-google-assistant-mic-950134/
60 - Marie Turcan, *Monsieur Cuisine Connect : micro caché, Android non sécurisé… les dessous du robot cuiseur de Lidl*, Numerama, juin 2019, https://www.numerama.com/tech/525214-monsieur-cuisine-connect-micro-cache-android-non-securise-les-dessous-du-robot-cuisine-de-lidl.html
61 - Surya Mattu et Kashmir Hill, *The House That Spied on Me*, Gizmodo, février 2018, https://gizmodo.com/the-house-that-spied-on-me-1822429852?rev=1518027891546
62 - Régis Chatellier, *L'espion qui me logeait : assistants vocaux et objets connectés dans la maison*, Linc.cnil.fr, avril 2018, https://linc.cnil.fr/fr/lespion-qui-me-logeait-assistants-vocaux-et-objets-connectes-dans-la-maison

Quelle répartition du marché ?

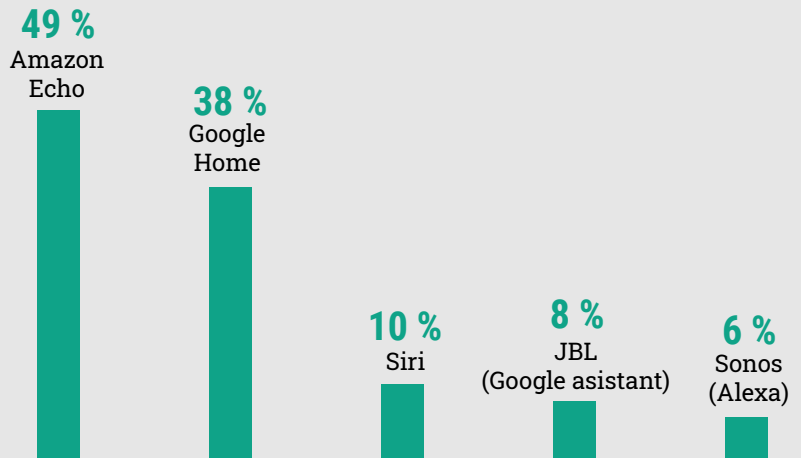
En pratique, le marché des assistants vocaux est stratégique pour les modèles de commerces en ligne et publicitaires. Ces derniers occupent une position hégémonique dans la vente d'assistants vocaux auprès du grand public alors qu'une pléiade de plus petits acteurs cherche à se positionner sur des marchés très ciblés et spécialisés. La bataille concurrentielle passe par une guerre des chiffres depuis le début d'année 2019. Si de son côté Amazon affirme avoir passé la barre des 100 millions de dispositifs connectés à Alexa[63], Google, rétorque avec 1 milliard d'objets qui intègrent Google Assistant[64]. Ces chiffres, qui ne font qu'évoluer, témoignent toutefois d'une tendance forte : si Amazon est derrière Google et Apple en termes d'appareils embarquant un assistant vocal, les chiffres de vente sur l'année 2019 indiquent que l'entreprise de e-commerce est le premier distributeur d'enceintes connectées dans le monde. Pour autant, ce constat est à contextualiser, comme le rappelle le rapport CSA-Hadopi : « Si Amazon domine largement le marché américain, les parts de marché des deux acteurs sont plus équilibrées au niveau mondial, témoignant de la force de Google sur les marchés d'export. En effet, Amazon domine le marché des enceintes connectées dans les marchés où son activité de e-commerce est bien implantée […] ». Si Facebook n'a jamais caché ses velléités de rentrer dans la course, la firme de Mark Zuckerberg n'a pas réussi pour l'instant à lancer son propre produit : son projet Aloha n'a pas (encore) vu le jour[65], et son outil de communication Portal se repose pour l'instant sur l'assistant d'Amazon Alexa[66]. Enfin, les constructeurs asiatiques tels que Baidu, Xiaomi, Alibaba ou encore Tencent font valoir de plus en plus leur volonté de se positionner sur le marché occidental et concurrencer les acteurs établis.

Concrètement, en France, selon une étude Statista de septembre 2019[67], sur ces quelques 1,7 millions d'enceintes connectées, l'estimation de la répartition du marché serait la suivante (pour les enceintes connectées les plus populaires) : 49 % d'Amazon Echo, 38 % de Google Home et 10 % de Siri. Ce à quoi il faut rajouter les enceintes JBL comprenant Google Assistant (8 %) et Sonos, qui embarquent Alexa (6 %). Des chiffres qui restent cependant à nuancer puisque l'enquête CSA-Hadopi montre une part plus importante des enceintes de Google par rapport à celles d'Amazon, mais qui dans l'ensemble illustrent bien la domination de ces deux acteurs (voir Figure 3).

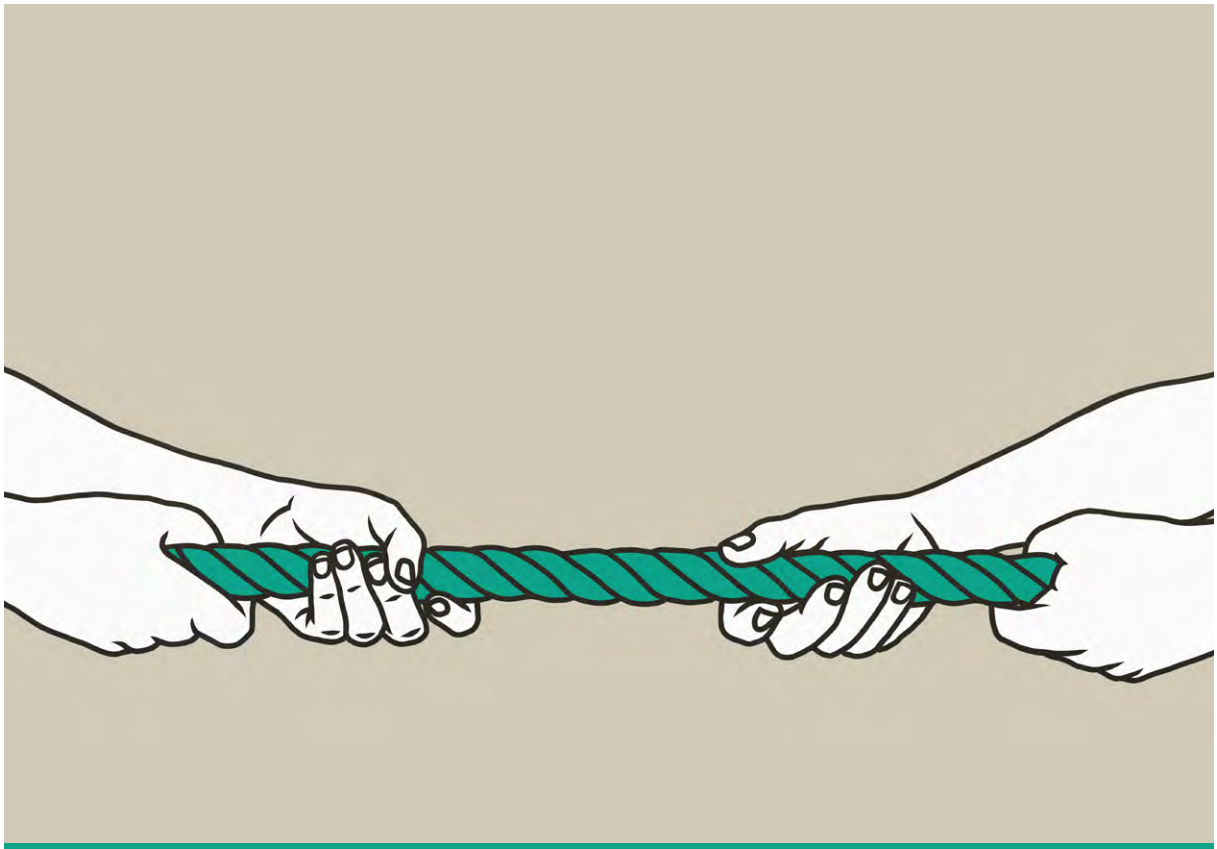
Figure 3

Répartition des enceintes équipées d'assistants vocaux en France en 2019

(source Statista).



63 - Dieter Bohn, *Amazon says 100 million Alexa devices have been sold — what's next?*, The Verge, janvier 2019, https://www.theverge.com/2019/1/4/18168565/amazon-alexa-devices-how-many-sold-number-100-million-dave-limp
64 - Scott Huffman, *Here's how the Google Assistant became more helpful in 2018*, Google keyword, janvier 2019, https://www.blog.google/products/assistant/heres-how-google-assistant-became-more-helpful-2018/
65 - Jean-Sébastien Zanchi, *Aloha : le projet secret d'assistant vocal de Facebook*, 01net, août 2018, https://www.01net.com/actualites/aloha-le-projet-secret-d-assistant-vocal-de-facebook-1509984.html
66 - Lucas Mediavilla, *Facebook développe son propre assistant vocal*, Les Echos, avril 2019, https://www.lesechos.fr/tech-medias/hightech/facebook-developpe-son-propre-assistant-vocal-1012806
67 - Statista, *Enceintes connectées : Amazon domine le marché français*, septembre 2019, https://fr.statista.com/infographie/19469/enceintes-connectees-les-plus-populaires-en-france/



LA VOIX SUR ÉCOUTE : MYTHES ET ENJEUX DES ASSISTANTS VOCAUX

Après avoir détaillé le fonctionnement, les usages et l'environnement des assistants vocaux, il convient désormais d'apprécier les conséquences que peuvent avoir leur utilisation. Les partis pris de conception, choix technologiques et plans de commercialisation posent des questions essentielles pour les utilisateurs : l'« objet vocal », par nature volatile et impalpable, trouve rigidité et consistance par l'usage de ces dispositifs.

MYTHES ET RÉALITÉS DES ASSISTANTS VOCAUX

De nombreuses idées reçues circulent sur les assistants vocaux et sur les capacités qui leurs sont prêtées. Retour sur cinq d'entre elles qui permettent de mieux comprendre la logique de fonctionnement de ces systèmes ainsi que les questions qu'ils posent pour leurs utilisateurs.

MYTHE n°1 : Ils écoutent en permanence

VRAI et FAUX (ils n'enregistrent pas tout !)

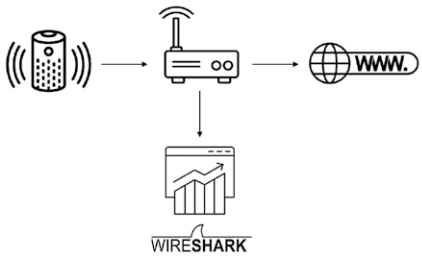
Comme précisé dans le Chapitre I.2 *Assistant vocal, qui es-tu ?*, et en particulier dans l'infographie détaillant le fonctionnement générique d'un assistant vocal (page 14), il convient pour l'utiliser que celui-ci soit « réveillé ». Cela signifie que l'assistant bascule dans un mode d'écoute actif afin de recevoir les ordres et commandes de son utilisateur. Si, dans de rares cas, ce réveil peut être réalisé suite à une action physique (par exemple en actionnant un bouton, en appuyant sur l'enceinte, etc.), la quasi-totalité des assistants vocaux du marché se base sur la détection d'un mot-clé (*keyword spotting*) pour passer en mode d'écoute actif (on parle également de mot d'activation ou en anglais *wake-up word / hot word*).

Pour cela, l'assistant s'appuie sur l'utilisation du microphone et de légères capacités calculatoires pour détecter si le mot-clé a été prononcé. Cette analyse, qui a lieu en permanence à partir du moment où l'assistant est en ser-

ZOOM SUR...

Les expérimentations du LINC

Depuis 2017, le Laboratoire d'Innovation Numérique de la CNIL (LINC) – la structure dédiée à l'expérimentation et à l'étude des tendances émergentes d'usage du numérique – s'est penché sur le sujet des assistants vocaux. Pour ce faire, différents tests et expérimentations ont été menés sur les équipements les plus répandus du marché.



Un des tests menés a en particulier visé à observer le trafic réseau montant issu d'un assistant vocal (en l'occurrence, les deux assistants les plus utilisés en France : Amazon Alexa et Google Assistant). Pour cela, a été utilisé un analyseur de paquets (Wireshark), placé en coupure de la connexion Internet (afin d'observer les caractéristiques des échanges entre l'assistant et les serveurs) comme indiqué dans le schéma ci-dessus. De plus, un environnement sonore proche de celui produit par un foyer a été simulé : l'assistant est placé dans une pièce de taille comparable à un salon, des conversations ont lieu et un téléviseur est susceptible d'être allumé par moment. Cette installation a permis de démontrer que, même si au cours de la période d'analyse aucun ordre spécifiquement destiné à l'assistant n'avait été formulé, de nombreuses activations avaient été enregistrées et étaient visibles dans l'historique d'utilisation. Par ailleurs, des chercheurs de Northeastern University et Imperial College London ont mené des expériences similaires et obtenu des résultats comparables[68].

68 - Daniel J. Dubois et al., *When speakers are all ears*, Smart speaker study, février 2020, https://moniotrlab.ccis.neu.edu/smart-speakers-study/

vice, est réalisée exclusivement localement et ce n'est que lorsque le mot-clé a été reconnu que les enregistrements audio sont traités pour interprétation et exécution de la commande, ce qui se traduit dans de nombreux cas par un envoi à des serveurs distants via Internet. La détection de mot-clé repose sur des techniques d'apprentissage automatique (*machine learning*). L'enjeu majeur du recours à de telles méthodes est que cette détection présente un caractère probabiliste. Ainsi, pour chaque mot énoncé, le système fournit un score de confiance quant au fait que le mot-clé ait été effectivement prononcé. Si ce score s'avère être supérieur à une valeur de seuil préalablement fixé, on considère que cela est bien le cas. Un tel système n'est donc pas exempt d'erreurs : dans certains cas l'activation peut ne pas être détectée alors que le mot-clé a pourtant été prononcé (faux rejet) et dans d'autres, une activation peut être détectée alors que l'utilisateur n'a pas prononcé le mot-clé (fausse acceptation).

En pratique, un compromis acceptable doit être trouvé entre ces deux types d'erreurs pour définir la valeur du seuil. Cependant, puisque la conséquence d'une fausse détection du mot-clé est la transmission d'enregistrements audio, des remontées d'information inopinées et non souhaitées sont susceptibles d'advenir. Bien souvent, les développeurs d'assistants vocaux mettant en œuvre des traitements distants utilisent un mécanisme à deux passes pour cette détection : une première embarquée localement au niveau de l'équipement et une seconde réalisée sur les serveurs distants où se dérouleront les traitements de données suivants (comme par exemple chez Apple[69] ou Google[70]). Dans ce cas, les développeurs ont tendance à mettre en place un premier seuil relativement bas pour favoriser l'expérience utilisateur et s'assurer que lorsqu'il prononce le mot-clé celui-ci est quasiment toujours reconnu – quitte à le « sur-détecter » – pour ensuite mettre en œuvre côté serveur une seconde passe de détection plus restrictive et consommatrice de ressources calculatoires.

Par conséquent, s'il n'y a pas a priori de volonté d'écoute en permanence de la part des fournisseurs d'assistants vocaux, la nature intrinsèquement statistique de la détection du mot-clé rend bien réel le risque de fausse activation (voir encadré ci-dessus). Il y a donc bien une écoute passive de la part des assistant vocaux (hors bouton d'activation paramétré), ce qui dans certains cas est susceptible d'avoir des conséquences néfastes pour les utilisateurs.

69 - Hey Siri: *An On-device DNN-powered Voice Trigger for Apple's Personal Assistant*, Apple Machine Learning Journal, octobre 2017, https://machinelearning.apple.com/2017/10/01/hey-siri.html

70 - Assaf Hurwitz et al., *Keyword Spotting for Google Assistant Using Contextual Speech Recognition*, IEEE Automatic Speech Recognition and Understanding Workshop, décembre 2017, https://research.google/pubs/pub46554/

Une question à… Julia Velkovska et Moustafa Zouinar

La promesse de ces dispositifs est la fluidité de l'échange avec l'utilisateur. Qu'en est-il en pratique selon vos observations ?

Tout d'abord la reconnaissance vocale, c'est-à-dire la retranscription de la parole humaine en texte exploitable par le système, n'est pas toujours efficace, y compris pour des requêtes simples comme demander la météo. Les utilisateurs doivent parfois répéter leurs énoncés plusieurs fois pour se faire comprendre, ce qui peut les conduire dans certains cas à abandonner l'usage du système. Dans les cas où ils persévèrent, ils s'engagent dans un véritable « travail de l'utilisateur », lié à la gestion de l'interaction et à son sens. Ce travail peut se traduire par une variété d'actions telles que reformuler les énoncés en les raccourcissant ou en les développant pour apporter des précisions, s'approcher de l'objet ou parler plus fort. Cet effort s'étend au-delà de la formulation des énoncés pour englober l'ensemble des activités accomplies par les personnes pour faire fonctionner le système, y compris celles visant à faire sens de ses réponses non pertinentes ou le travail d'apprentissage lié à la « structure » interactionnelle imposée (activer le système puis parler au bon moment, c'est-à-dire lorsqu'il est « en écoute »). Ce travail de l'utilisateur est un aspect majeur des usages actuels des assistants vocaux et il est très important de le décrire, de le comprendre et d'en tenir compte lorsqu'on pense les conséquences sociales de la diffusion de ces technologies à l'époque où elles sortent justement des laboratoires pour prendre place au cœur des foyers..

Julia Velkovska et Moustafa Zouinar sont respectivement sociologue et ergonome au Laboratoire SENSE, Orange Labs

> Entretien intégral à retrouver sur LINC
Julia Velkovska et Moustafa Zouinar : Assistants vocaux : un véritable fossé entre les discours promotionnels et la réalité des usages, Linc.cnil.fr, avril 2018, https://linc.cnil.fr/fr/julia-velkovska-et-moustafa-zouinar-assistants-vocaux-un-veritable-fosse-entre-les-discours

MYTHE n°2 : Ils nous comprennent parfaitement

FAUX

S'ils mettent en œuvre de multiples techniques d'intelligence artificielle (IA), prêter des capacités de production d'un comportement intelligent, de modélisation d'idées abstraites, mais aussi de conscience et de sentiments à un assistant vocal tient toujours du domaine de la science-fiction. Si certains d'entre eux dits « généralistes » sont susceptibles d'être adaptés pour répondre à tout type de sollicitation, ils ne peuvent pas pour autant être qualifiés d'« IA forte ». Les assistants vocaux sont un avatar de ce qu'on a plutôt coutume d'appeler « IA faible », c'est-à-dire des systèmes de plus en plus autonomes mettant en œuvre des algorithmes capables de résoudre des problèmes d'un type donné. A contrario, par rapport à la notion d'« IA forte », la machine simule l'intelligence et semble agir comme si elle était intelligente, comme par exemple en imitant le comportement d'une personne face à une autre lors d'un dialogue.

Si aujourd'hui les assistants vocaux fournissent des résultats globalement satisfaisants, il convient toutefois de noter qu'on reste loin d'échanges aussi fluides et naturels que vantés par les promoteurs de ces technologies. Ainsi, comme le rappellent Julia Velkovska et Moustafa Zouinar (voir encadré), il est nécessaire pour l'utilisateur de réaliser un véritable « travail ». En effet, lorsque deux personnes parlent ensemble, on observe un phénomène de convergence interactionnelle au cours duquel chacune prend les habitudes discursives de l'autre de façon à créer un espace commun qui va favoriser la compréhension mutuelle. Dans le cas d'un échange avec un assistant vocal (ou toute autre machine parlante), c'est l'utilisateur qui, seul de son côté, opère cette convergence afin de s'adapter aux capacités de compréhension de la machine.

Enfin, la raison pour laquelle les utilisateurs sont parfois choqués des dérapages et défaillances des assistants vocaux tient peut-être de la théorie de « la vallée de l'étrange » (*the uncanny valley*) inventée dans les années 1970 par le roboticien japonais Masahiro Mori. D'après celle-ci, plus un robot est similaire à un être humain et plus ses différences apparaissent monstrueuses. Transposée aux cas des assistants vocaux, cela implique que plus nombreuses sont les questions auxquelles un assistant est à même de répondre, plus l'utilisateur est frappé lorsqu'il faillit. Il apparaît donc nécessaire de délimiter précisément le domaine d'intervention d'un assistant au moment de sa création, sous peine d'engendrer une vraie désillusion chez ses utilisateurs.

MYTHE n°3 : Ils utilisent nos données pour mieux nous profiler

VRAI ET FAUX

Nous avons vu dans le Chapitre I.4 *Quelle(s) stratégie(s) pour les concepteurs d'assistants vocaux ?* que les concepteurs d'assistants vocaux se fondaient sur différents modèles d'affaires. Dans certains d'entre eux, il s'agit de fournir un service à un utilisateur en ayant comme unique but de permettre à l'utilisateur une nouvelle modalité d'interaction avec un équipement passant par la voix. C'est notamment le cas des assistants vocaux intégrés en marque blanche à la demande de certains constructeurs de téléviseurs, d'aspirateurs, etc. Les interactions de l'utilisateur avec son assistant n'alimentent donc généralement pas, dans ces cas, un écosystème de la donnée.

Toutefois, pour les principaux fournisseurs d'assistants, en particulier les grand acteurs du numérique comme Google et Amazon dont l'activité repose sur l'exploitation des données à caractère personnel de leurs utilisateurs, il ne s'agit là que d'un nouveau vecteur de collecte. Principalement destinés au domicile pour contrôler des objets connectés, des services de divertissement ou des applications domotiques, les appareils dotés d'un assistant vocal se retrouvent au cœur de la vie du foyer. Bien souvent, il est nécessaire de créer un compte utilisateur pour pouvoir profiter pleinement des options proposées par l'assistant. Les grands constructeurs proposent ainsi de lier directement les échanges avec leurs assistants à des comptes utilisateurs déjà existants pour utiliser leurs produits (email, agenda, magasin en ligne, etc.). Ainsi, le profil des utilisateurs se trouve alimenté (dans le cas d'un nouveau compte) et complété (dans le cas d'un compte existant) par les différentes interactions de l'utilisateur avec l'assistant : habitudes de vie (heures de lever et de coucher), réglages du chauffage, goûts culturels, achats passés, centres d'intérêt, etc. En permettant ainsi d'enrichir les informations déjà existantes sur un utilisateur, il est possible de mettre en œuvre des campagnes publicitaires plus ciblées, d'offrir des propositions commerciales plus fines et corrélées aux informations collectées, etc. Un tel positionnement fait cependant apparaître une tension entre la mise en œuvre d'un modèle de profilage centré sur l'individu et le caractère potentiellement très collectif de l'assistant, par exemple lorsque déployé au sein d'un foyer. Si l'assistant vocal peut permettre de collecter des informations nombreuses sur le collectif au sein duquel il est déployé, une multiplication des utilisateurs peut également être source de confusion pour les modèles personnalisés, avec par exemple des propositions com-

merciales ou annonces publicitaires de biens ou services qui pourraient être faites suite aux interactions d'autres personnes avec l'assistant (conjoint, enfants, etc.).

Pour les acteurs mettant en œuvre de telles stratégies d'alimentation de profils, il est essentiel de promouvoir leur rôle d'intermédiaire et d'orchestrateur auprès de professionnels tiers (fournisseur d'objets connectés, vendeurs en ligne, organismes de transports, etc.) afin de les convaincre du bien-fondé d'être présent sur leur plateforme en développant une application ad-hoc. En effet, plus l'offre sur ces plateformes est importante, plus les utilisateurs sont sujets à interagir fréquemment avec l'assistant et plus les informations détenues sur eux sont fines. De ce fait, le nombre d'équipements adressables, c'est-à-dire équipés de l'assistant, est un argument de vente décisif pour encourager les développeurs d'applications tierces !

Enfin, si le profilage est pour l'instant uniquement réalisé à partir d'informations inférées des commandes passées, la question d'un élargissement de la portée de la collecte peut se poser, sans préjuger de la licéité de telles pratiques. Par exemple, des informations relatives au bruit ambiant obtenues par analyse de scènes sonores (chaînes de télévision regardées, enfant en bas âge pleurant en arrière-plan, etc.) ou à l'état émotionnel ou de santé par analyse du signal vocal sont susceptibles d'intéresser les développeurs d'assistants. À titre d'exemple, la société Amazon, a ainsi déposé en 2017 un brevet visant à catégoriser les états émotionnels de ses utilisateurs[71].

MYTHE n°4 : Ils sont une interface très prisée des enfants

VRAI

Comme vu dans le Chapitre précédent, les enfants ont une forte appétence pour ces dispositifs. Ils leur permettent – du moins en théorie – d'accéder à des ressources généralement hors de leur portée comme par exemple une chaîne hi-fi et ainsi de diffuser la musique de leur choix. Toutefois, les énonciations tortueuses et hésitantes des enfants sont bien souvent mal retranscrites et donc également mal interprétées ce qui occasionne chez eux une frustration importante et les oblige également à effectuer un « travail » pour s'adapter aux capacités de l'assistant (voir Mythe n°2).

En 2014, Johan Schalkwyk, responsable des programmes relatifs au traitement automatique de la parole chez Google, annonçait que « nos enfants et petits-enfants ne comprendront pas qu'on ait un jour pu interagir avec un ordinateur autrement que par la voix ». Ainsi très tôt, les géants du numérique ont identifié la voix comme l'interface homme-machine (IHM) du futur, accessible facilement par tous et partout. Les campagnes publicitaires des grands acteurs insistent d'ailleurs sur cet aspect, l'assistant étant censé permettre des interactions faciles et fluides, y compris aux publics les plus éloignés des codes et pratiques numériques, à savoir les jeunes enfants et les personnes âgées. La communication de Facebook sur son outil Portal l'illustre bien : il s'agit de permettre à toute la famille de rester en contact, des enfants aux grands-parents. Le jeune public se voit également offrir des contenus spécifiquement produits pour lui : applications de blagues, de jeux, d'histoires lues, etc., dont de nombreuses cachent en vérité une finalité commerciale puisque développées par telle ou telle marque. Qui plus est par la manipulation de cette interface et grammaire d'usage, se façonne également « l'éducation » de futurs consommateurs.

Si les assistants vocaux peuvent présenter une alternative aux écrans, le fait de laisser un tel dispositif librement accessible à des enfants suscite toutefois des questions. De nombreux exemples existent illustrant les risques potentiels : réalisation d'achats en ligne[72], accès à des contenus pour adultes[73], etc.

De façon générale, et comme pour tous les objets connectés (dont les jouets), il est essentiel de mettre en œuvre les bonnes pratiques élémentaires de protection. La CNIL a eu l'occasion de communiquer à ce sujet[74]. Appliqué aux cas des assistants vocaux, il convient en particulier de s'assurer de l'utilisation de techniques de filtrage parental et d'encadrer les interactions des enfants. À titre d'exemple, la société Mattel a abandonné l'idée de lancer l'assistant vocal Aristotle qui devait équiper ses produits, en offrant notamment une plateforme sur laquelle les parents pouvaient écouter, voire réécouter les conversations que les enfants avaient avec leurs jouets[75]. Une pratique qui peut potentiellement nuire à la vie privée de l'enfant ou affecter le rapport de confiance qu'il entretient avec ses parents.

71 - Jon Brodkin, *Amazon patents Alexa tech to tell if you're sick, depressed and sell you meds*, Ars Technica, novembre 2018, https://arstechnica.com/gadgets/2018/10/amazon-patents-alexa-tech-to-tell-if-youre-sick-depressed-and-sell-you-meds/
72 - Karma Allen, *6-Year-Old Mistakenly Orders Dollhouse, Cookies Worth $162 While Chatting With Amazon Echo*, ABC News, janvier 2017, https://abcnews.go.com/Technology/year-mistakenly-orders-162-worth-treatschatting-amazon/story?id=44577327
73 - Post staff report, *Boy requests song from Amazon Alexa, but gets porn instead*, New York Post, décembre 2016, https://nypost.com/2016/12/30/toddler-asks-amazons-alexa-to-play-song-but-gets-porn-instead/
74 - CNIL, ***Jouets connectés : quels conseils pour les sécuriser ?***, https://www.cnil.fr/fr/jouets-connectes-quels-conseils-pour-les-securiser
75 - Rachel Rabkin Peachman, *Mattel Pulls Aristotle Children's Device After Privacy Concerns*, The New York Times, octobre 2017, https://www.nytimes.com/2017/10/05/well/family/mattel-aristotle-privacy.html

MYTHE n°5 : Ils sont piratables

PLUTÔT VRAI

Incarnés dans des équipements lisses et épurés, il est fréquent d'oublier la nature purement logicielle des assistants vocaux. Pourtant, intégrés dans des équipements du quotidien, ceux-ci appartiennent par nature à la galaxie de l'Internet des objets (*Internet of Things* ou *IoT*). À ce titre, ils sont tout aussi sujets à des attaques informatiques.

Le nombre d'attaques effectives véritablement recensées à l'encontre des assistants vocaux reste encore aujourd'hui assez faible. Les attaques connues et reportées depuis quelques années semblent principalement avoir été le fait de chercheurs en sécurité (voir encadré). Toutefois, comme pour tous les équipements de l'Internet des objets, il est à prévoir qu'avec le développement des services proposés, les pirates trouveront de plus en plus d'intérêt à accéder à ces appareils de façon illégitime, que ce soit pour en prendre le contrôle ou pour accéder aux données qui y transitent. Qui plus est, avec l'ambition de faire des assistants vocaux le centre névralgique du foyer connecté (*smart hub*), ceux-ci peuvent devenir un point central de vulnérabilité du système d'information de la maison « intelligente ».

ZOOM SUR...

Panorama des défaillances et attaques

Les premières attaques perpétuées (et documentées) à l'encontre des assistants vocaux semblent avoir été celles mises en œuvre par la société Burger King[76] et la série South Park[77] en 2017. Dans les deux cas, des contenus audiovisuels contenant les mots-clés activant les principaux assistants (« Hey Alexa », « Ok Google ») suivi de commande visant à transmettre des informations publicitaires ou farfelues ont été diffusés. Dans ces deux exemples, les attaquants se reposent sur deux faits : 1) un assistant vocal peut-être activé par toute personne à distance d'écoute et 2) les enceintes connectées se situent bien souvent dans le salon, à l'endroit même où se trouve le téléviseur du foyer. Cette attaque, certes rudimentaire, a obligé les concepteurs d'assistants vocaux à déployer des stratégies de contournement empêchant l'activation soit en réalisant une signature audio du contenu en vue de l'identifier et le bloquer (ce qui implique une exposition préalable), soit en effectuant une analyse en continu des activations des assistants vocaux et en bloquant l'exécution de la commande lorsqu'un volume trop important d'activations simultanées est observé.

Depuis, la communauté de recherche en sécurité informatique s'est largement saisie du nouvel objet d'étude que constituent les assistants vocaux. Ainsi de très nombreuses attaques ont démontré que les plus célèbres assistants vocaux étaient susceptibles d'être activés et utilisés à l'insu de leurs propriétaires. Le premier grand type d'attaque, comme la *Dolphin attack*[78] (mais également les *Backdoor attack*[79] et *LipRead attack*[80]) s'appuie sur le fait que les microphones des assistants vocaux sont sensibles aux fréquences hautes situées hors du spectre audible (supérieur à 20 000 hertz). Il est ainsi possible, à l'aide d'équipements rudimentaires (smartphone, amplificateur et transducteur pour ultrasons) de passer à un assistant des commandes imperceptibles. La seconde famille d'attaque se sert du phénomène psychoacoustique de « masquage » pour cacher des commandes à destination d'assistants vocaux dans d'autres signaux audio (comme par exemple le pépiement des oiseaux dans la *Chirping birds attack*[81]).

(suite P36...)

76 - Antoine Boudet, *Burger King détourne Google Home pour faire sa publicité, Google riposte*, Numérama, avril 2017, https://www.numerama.com/tech/249062-burger-king-detourne-google-home-pour-faire-sa-publicite-google-riposte.html
77 - Vincent Tanguy, *Quand South Park rend fous Alexa et Google Home*, Sciences et Avenir, septembre 2017, https://www.sciencesetavenir.fr/high-tech/intelligence-artificielle/quand-south-park-rend-fous-alexa-et-google-home_116434
78 - Guoming Zhang et al., *DolphinAttack: Inaudible Voice Commands*, ACM Conference on Computer and Communications Security, novembre 2017, https://acmccs.github.io/papers/p103-zhangAemb.pdf
79 - Nirupam Roy et al., *BackDoor: Making Microphones Hear Inaudible Sounds*, MobiSys 2017, juin 2017, https://synrg.csl.illinois.edu/papers/backdoor_mobisys17.pdf
80 - Nirupam Roy et al., *Inaudible Voice Commands: The Long-Range Attack and Defense*, USENIX Symposium on Networked Systems Design and Implementation, avril 2018, https://synrg.csl.illinois.edu/papers/lipread_nsdi18.pdf
81 - Lea Schönherr et al., *Adversarial Attacks Against ASR Systems via Psychoacoustic Hiding*, Network and Distributed System Security Symposium, février 2020, https://adversarial-attacks.net/

(...suite) De façon plus générale, on parle ici d'attaques adversaires qui permettent, comme démontré dans plusieurs travaux[82][83], de masquer n'importe quelle commande dans un enregistrement audio sans que celui-ci soit décelable sans une analyse par ordinateur. D'autres attaques, plus surprenantes encore, ont également été menées. La *Surfing attack*[84] permet par exemple d'injecter des commandes ultrasons en faisant vibrer la surface sur laquelle est posé le dispositif équipé d'un assistant vocal (par exemple une table). Enfin, par une *Laser attack*[85], un pirate peut envoyer des commandes inaudibles et invisibles en pointant un laser sur la membrane du haut-parleur, l'attaque pouvant être menée à une centaine de mètres de distance !

Plusieurs recherches ont également montré que les assistants vocaux pouvaient être utilisés à des fins d'hameçonnage (le terme phishing transformé en vishing dans le cas de la voix). Ainsi, la technique du Skill squatting[86], permet à un pirate de créer une application homophone d'une application légitime en se basant sur les limitations du système de transcription automatique de la parole (*speech to text*). À titre d'exemple, les chercheurs développent une application nommée « *Am Express* » qui peut être utilisée pour détourner les requêtes faites à l'application « Amex », d'American Express, et mettre en œuvre des interactions avec un utilisateur afin d'accéder à ses informations personnelles et financières. Enfin, d'autres techniques d'hameçonnage et d'écoutes illicites sont également mises en œuvre[87]. Celles-ci s'appuient en particulier sur le fait de faire croire à l'utilisateur que l'interaction avec l'application qu'il a appelée est terminée alors qu'elle se poursuit, par exemple, en demandant à l'assistant de jouer une séquence imprononçable qui rend l'assistant silencieux alors que l'application, elle, est encore en cours d'exécution.

Il faut également rappeler que les caractéristiques spécifiques de l'assistant les rendent vulnérables sans pour autant passer par des moyens nécessitant des connaissances techniques élevées. En effet, le rayon d'activation d'un assistant est autour de cinq mètres : être à cette portée peut permettre à une personne mal intentionnée d'utiliser l'assistant (ouvrir une porte, récupérer des informations, etc.). De la même manière, l'assistant fonctionne sur un principe de capteur (un micro), qu'il est tout à fait possible de brouiller : en créant un bruit ambiant par exemple, qu'il soit audible ou pas par l'homme (bruit blanc ou *white noise* qui couvre toutes les fréquences de façon uniforme). Dès lors, il serait possible d'empêcher le fonctionnement de certains dispositifs se reposant sur la seule reconnaissance vocale. Certains, comme le projet Alias[88], utilisent d'ailleurs ce principe pour lutter contre les activations intempestives des assistants vocaux. Il s'agit dans ce cas d'une surcouche protectrice se plaçant au-dessus d'une enceinte équipée d'un assistant vocal. Celle-ci diffuse en permanence du bruit blanc sauf quand un mot-clé, défini préalablement par l'utilisateur est prononcé. À partir de cet instant, l'émission du bruit parasite s'interrompt et l'assistant vocal devient alors utilisable de façon traditionnelle.

Outre les attaques, informatiques ou non, des défaillances sont également susceptibles d'advenir et plusieurs d'entre elles ont déjà été rendues publiques. Les appareils Google Home mini – dont une particularité était de pouvoir être activé par une simple pression du doigt – ont ainsi été mis en cause en 2017 lorsqu'un journaliste s'est aperçu en consultant son historique d'utilisation que le sien était activé plusieurs milliers de fois par jour[89]. En effet, l'appareil détectait des « appuis fantômes » et était de ce fait activé en permanence. Enfin, autre raté notable, à l'occasion d'une demande de droit d'accès faite auprès d'Amazon, un utilisateur allemand a obtenu l'historique des interactions avec Alexa d'un autre détenteur de compte Amazon. Un travail d'investigation mené par des journalistes a permis de retrouver l'identité de la personne à qui ces données appartenaient et celle-ci a par la suite porté plainte[90].

82 - Tavish Vaidya et al., *Cocaine Noodles: Exploiting the Gap between Human and Machine Speech Recognition*, USENIX Workshop on Offensive Technologies, août 2015, https://www.usenix.org/node/191969
83 - Nicholas Carlini et David Wagner, *Audio Adversarial Examples: Targeted Attacks on Speech-to-Text*, IEEE Security and Privacy Workshops (SPW), mai 2018, https://arxiv.org/pdf/1801.01944.pdf
84 - Qiben Yan et al., *SurfingAttack: Interactive Hidden Attack on Voice Assistants Using Ultrasonic Guided Waves*, Network and Distributed System Security Symposium, février 2020, https://surfingattack.github.io/
85 - Takeshi Sugawara et al., *Light Commands: Laser-Based Audio Injection on Voice-Controllable Systems*, novembre 2019, https://lightcommands.com/
86 - Deepak Kumar et al., *Skill Squatting Attacks on Amazon Alexa*, USENIX Security Symposium, août 2018, https://www.usenix.org/conference/usenixsecurity18/presentation/kumar
87 - Security Research Labs, Smart Spies: *Alexa and Google Home expose users to vishing and eavesdropping*, novembrenovembre 2019, https://srlabs.de/bites/smart-spies/
88 - Bjørn Karmann, *Project Alias*, 2018, http://bjoernkarmann.dk/project_alias
89 - Artem Russakovskii, *Google is permanently nerfing all Home Minis because mine spied on everything I said 24/7*, Android Police, octobre 2017, https://www.androidpolice.com/2017/10/10/google-nerfing-home-minis-mine-spied-everything-said-247/
90 - Holger Bleich, *Amazon reveals private Alexa voice data files*, Heise online, décembre 2018, https://www.heise.de/newsticker/meldung/Amazon-reveals-private-voice-data-files-4256015.html

QUELS ENJEUX POUR LES ASSISTANTS VOCAUX ?

Des enjeux éthiques...

L'apparition de ces nouveaux dispositifs d'interaction sur des smartphones comme au sein des foyers soulève des questions éthiques. Puisqu'ils modifient nos façons de rechercher et de trouver l'information, de vivre notre maison (et au sein de celle-ci via la domotique), d'interagir avec les services personnels (mail, agenda ou même compte bancaire), ils touchent à notre environnement au sens large. Dès lors, il est pertinent de se pencher sur ces changements et leurs conséquences tant dans le fonctionnement que dans les usages.

1) De l'écran à la voix, la création d'une nouvelle grammaire d'usage

Les assistants vocaux proposent un changement de paradigme majeur en transformant le rapport aux outils numériques, à travers le passage du visuel au vocal. Les interfaces graphiques s'effacent peu à peu, et ces dispositifs, par souci de fluidification, laissent le champ libre à la voix. L'écran demeure un support secondaire, à travers des comptes et applications « compagnons », ceux-ci permettant notamment de configurer le système et de suivre son activité. Toutefois, le but est de s'affranchir au maximum de tout support visuel ainsi que des boutons apparents. Si, comme on l'a vu précédemment, cette nouvelle interface homme-machine présente de grands avantages (voir Chapitre I.3 *Quels usages pour les assistants vocaux ?*), elle pose également des questions. Sans écran, difficile d'avoir un aperçu des traces enregistrées, ni de juger de la pertinence des suggestions, d'en savoir plus ou de tracer les informations qui sont fournies. Alors que de nombreux acteurs du numérique souhaitent « faire disparaître la technologie », un vrai risque d'opacité de ces systèmes existe et les modalités d'information et de consentement de l'individu doivent être particulièrement réfléchies. Peut-on être bien informé vocalement ? Premier lieu de médiation entre l'individu et l'information et entre l'individu et ses droits, les interfaces nécessitent une construction solide et claire permettant à tout un chacun de comprendre les enjeux de l'utilisation d'un service ou produit. Depuis plusieurs années, la CNIL mène des réflexions et travaux sur le design de celles-ci, débouchant sur la publication en janvier 2019 du Cahier IP n°6 *La forme des choix*[93], ainsi qu'un site web dédié Données & Design[94].

ZOOM SUR...

Le CNPEN et les enjeux éthiques des agents conversationnels

Le Comité national pilote d'éthique du numérique (CNPEN) a été créé en décembre 2019 à la demande du Premier ministre[91]. Constitué de 27 membres, dont un représentant de la CNIL, ce comité réunit des spécialistes du numérique, des philosophes, des médecins, des juristes et des membres de la société civile. L'une des trois saisines soumises par le Premier ministre au CNPEN concerne les enjeux éthiques des agents conversationnels, incluant notamment les assistants vocaux. Dans ce cadre, et pour préparer ses recommandations à l'attention des concepteurs ainsi que des utilisateurs des agents conversationnels, le CNPEN a lancé un appel à contributions ouvert jusqu'au 30 septembre 2020[92]. Cet appel vise à permettre une expression des parties prenantes et du public sur les questions éthiques liées à ces « *chatbots* » de plus en plus présents dans nos vies. Les enjeux éthiques présentés ici s'articulent avec les réflexions éthiques menées par le CNPEN qui prolongent les travaux initiés par la CERNA, la Commission de réflexion sur l'éthique de la recherche en science et technologies du numérique de l'alliance Allistene.

91 - Comité Consultatif National d'Éthique, Création du Comité Pilote d'Éthique du Numérique, décembre 2019, https://www.ccne-ethique.fr/fr/actualites/creation-du-comite-pilote-dethique-du-numerique
92 - CNPEN, Les enjeux éthiques des agents conversationnels, juin 2020, https://www.ccne-ethique.fr/fr/actualites/cnpen-les-enjeux-ethiques-des-agents-conversationnels
93 - LINC, Cahier IP6 : *La forme des choix*, janvier 2019, https://linc.cnil.fr/fr/cahier-ip6-la-forme-des-choix-0
94 - CNIL, Données & Design : une nouvelle plateforme pour la communauté des designers autour du RGPD, https://www.cnil.fr/fr/donnees-design-une-nouvelle-plateforme-pour-la-communaute-des-designers-autour-du-rgpd

2) Des multiples pages de résultats à la réponse unique

Le rapport à l'assistant vocal engendre également un changement de paradigme dans la manière d'appréhender la recherche d'information en ligne. Contrairement au fonctionnement classique d'un moteur de recherche, l'assistant va choisir une seule réponse à la question posée et la délivrer telle quelle. Il y aura donc une réponse unique à une question, là où auparavant il était possible de naviguer dans les pages de résultats. On passe ainsi d'un moteur de recherche à un moteur de réponse. Le fait de ne fournir qu'une réponse unique soulève des interrogations quant au choix des sources de résultats. Pour des questions simples, du moins en apparence, le problème peut ne pas se poser (météo, encyclopédie, etc.). Dans d'autres cas, lorsqu'il s'agit d'actualités par exemple politiques, le choix des sources à privilégier se pose. Comment assurer la neutralité de celles-ci ? Et comment en assurer la fiabilité ? Les réponses varient-elles selon le profil de l'individu ? Cela est-il souhaitable ?

L'exemple de Wikipédia est particulièrement parlant. Les contenus proposés par les rédacteurs sont corrigés, vérifiés et modérés par la communauté, générant un effet de neutralité. Mais par ce fonctionnement ouvert, Wikipédia peut être temporairement victime de détournement. La modification d'un contenu peut alors contribuer à la diffusion de propos biaisés ou publicitaires, comme ceux que la marque The North Face avait publiés pour mettre en avant ses produits en 2019 (ce qui lui avait valu de vives critiques, l'obligeant à produire un billet d'excuse[95]). Des cas de détournement contenant des propos agressifs ont également été documentés. À titre d'exemple, une version d'un article Wikipédia sur le cycle cardiaque conseillant alors de « se poignarder dans le cœur », car l'activité cardiaque contribuait à « l'appauvrissement de la planète », a pu être lu par des assistants vocaux[96]. En l'occurrence, il s'agissait d'un canular fait par un contributeur de Wikipédia mal intentionné, et dont les modifications ont rapidement été écartées par la communauté.

3) De l'interface homme-machine à la relation homme-androïde ?

a) Une tendance à l'anthropomorphisation

Les voix de synthèse n'ont pas attendu le développement des assistants vocaux pour s'installer dans la vie des individus et populations, en particulier dans l'espace urbain (dans les gares, transports en commun, au passage piétons, etc.). Dans un but d'adhésion plus forte des utilisateurs, les concepteurs ont fait le choix de passer par une personnification, préférée à un modèle artificiel et mécanique. Dans le cas spécifique des assistants vocaux, ce choix a été fait dans l'optique d'un engagement plus important. En reproduisant les conditions d'un dialogue en langage naturel, les interactions avec l'outil sont facilitées et, ainsi, plus nombreuses. Les individus peuvent directement poser leurs questions à l'assistant de la même manière qu'ils interrogeraient une autre personne, le mot-clé en plus. De son côté, l'assistant prend le temps de répondre selon des codes de langage établis (sujet – verbe – complément), et pas seulement en délivrant une information brute. En lui donnant un nom humanisé (Siri, Alexa, Cortana, Célia, etc.), des caractéristiques et une histoire, en lui faisant faire des blagues, à travers la richesse de ses réponses, les concepteurs ont ainsi attaché une fiction et une personnalité à leur voix. Le but est de dépasser le simple outil pour devenir un interlocuteur et un compagnon. L'utilisateur est d'ailleurs amené à paramétrer son dispositif, notamment en ce qui concerne les caractéristiques de la voix avec laquelle il va interagir (genre, accents, etc.). La chercheuse en sciences de l'information et de la communication Clotilde Chevet précise cet aspect dans ses travaux en revenant sur le concept de syndrome Ikea. Ce dernier, développé par les chercheurs Shyam Sundar et Yuan Sun, montrerait en effet un attachement plus fort à un objet si l'on a contribué à le construire[97].

b) Des caractéristiques genrées de l'assistant

Les caractéristiques des assistant(e)s vocaux (noms, voix) peuvent contribuer à la reproduction d'une représentation inégalitaire de la place des femmes dans la société, à commencer par le fait de les placer dans le rôle de l'assistante.En 2019, l'UNESCO a publié un rapport consacré à cette question, intitulé « I'd blush if I could » (soit « Je rougirais si je le pouvais »), en référence à ce que pouvait répondre Siri lorsque l'assistant vocal était insultée de manière misogyne[98]. La réponse de l'assistant avait provoqué de nombreuses critiques et des modifications avaient été apportées : « Je ne sais pas comment répondre à cela ». Le constat de la reproduction de stéréotypes de genres à travers les interfaces vocales a mené au développement de projets d'assistants vocaux neutres et non-genrés tels que Q[99].

c) La confusion entre agent virtuel et interlocuteur humain

La relation à ces objets n'est également pas sans poser un certain nombre de questions. Comment doit-on considérer

95 - LeBrief, *The North Face détourne Wikipédia pour placer ses produits, puis s'excuse*, NextInpact, mai 2019, https://www.nextinpact.com/brief/-the-north-face-detourne-wikipedia-pour-placer-ses-produits--puis-s-excuse---8860.htm
96 - James Crowley, *Woman Says Amazon's Alexa Told Her To Stab Herself In The Heart For 'The Greater Good'*, Newsweek, décembre 2019 https://www.newsweek.com/amazon-echo-tells-uk-woman-stab-herself-1479074
97 - Clotilde Chevet, *La voix de synthèse : de la communication de masse à l'interaction homme-machine.* Dialogue avec le monde, Communication & langages 2017/3 (N° 193), https://www.cairn.info/revue-communication-et-langages1-2017-3-page-63.htm
98 - UNESCO, EQUALS Skills *Coalition, I'd blush if I could: closing gender divides in digital skills through education*, 2019, https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1
99 - https://www.genderlessvoice.com/

ces nouveaux dispositifs au sein de l'espace domestique ? Quelle place doit-on leur laisser dans les foyers ? Et au sein de la famille ? Comme vu précédemment, les plus jeunes représentent une part importante des utilisateurs des enceintes connectées équipées d'assistants vocaux, ce qui pose notamment des questions en termes d'éducation. La journaliste et membre du comité de la prospective de la CNIL, Titiou Lecoq, a pointé les questions qui se posent dans la relation de l'enfant à l'assistant[100]. En effet, faut-il lui dire merci ? Comment faire comprendre à un enfant la différence qui existe entre une personne réelle (humaine) et une intelligence artificielle (IA) ? Comment résister à la possibilité de donner un ordre à une voix d'adulte qui ne pourra jamais vous punir ?

4) Les assistants vocaux sont des systèmes d'IA comme les autres

a) La présence de biais

Les assistants vocaux utilisent des technologies d'intelligence artificielle. Ils se reposent sur des méthodes d'apprentissage automatique (*machine learning*) pour la réalisation de très nombreuses tâches : détection du mot-clé, transcription automatique de la parole, compréhension du langage, synthèse de la parole, etc. La CNIL a publié en 2017 un rapport éthique intitulé *Comment permettre à l'homme de garder la main*[101] qui questionne les enjeux éthiques des algorithmes. La question des biais que des systèmes apprenant sont susceptibles de comporter y est développée ainsi que celle des risques qu'ils impliquent pour les utilisateurs. En effet, ces systèmes informatiques apprennent à partir d'ensembles de données qui ont été collectés, sélectionnés, étiquetés, etc. mais qui peuvent contenir des biais. Ces surreprésentations ou au contraire sous-représentations de certaines populations ou caractéristiques peuvent infléchir l'apprentissage de l'IA et répercuter par la suite cette erreur ou un mauvais paramétrage dans ses calculs, et donc dans sa manière de fonctionner[102]. La qualité des données joue donc un rôle majeur dans la finesse et la précision de l'apprentissage, au même titre que leur quantité. Dans le cadre des assistants vocaux, les corpus de textes utilisés peuvent statistiquement faire remonter plus fréquemment des pronoms de genre féminin que masculin, les voix servant à l'entrainement peuvent être principalement celles d'adultes alors que le système est prévu pour pouvoir interagir avec des enfants, etc. Qui plus est les différents timbres de voix, types d'élocution, accents, langages, etc. peuvent introduire des sources d'erreurs spécifiques à la reconnaissance vocale. S'il est possible d'anticiper ces biais et d'injecter des correctifs dans les systèmes, tous les biais ne sont pas forcément connus à l'avance.

b) Le recours à des travailleurs du clic

Que ce soit pour qualifier la base de données d'apprentissage ou pour corriger les erreurs commises lorsque l'algorithme est déployé, l'apprentissage et l'entrainement des systèmes d'intelligence artificielle nécessitent obligatoirement une intervention humaine. Cette partie du travail, qu'on qualifie de travail du clic (*digital labor* en anglais), soulève des critiques. Le sociologue Antonio Casilli décrit dans ses travaux l'externalisation de ces micro-tâches à des personnes peu qualifiées et peu payées, parfois dans des régions du monde où les salaires sont peu élevés, afin de réduire les coûts[103]. Cela soulève des questions tant sur les conditions de travail, qu'en termes de sécurité. En effet, on observe dans de nombreux cas que les données, matière première de ces micro-tâches, peuvent circuler entre concepteurs de systèmes d'intelligence artificielle et sous-traitants sans que les garanties nécessaires ne soient mises en œuvre[104].

...et plus spécifiquement, des enjeux pour la vie privée

Les interactions avec les assistants alimentent une collecte d'informations relatives à la vie quotidienne et à l'intimité. Qui plus est cette collecte est amplifiée par l'utilisation d'applications tierces, ces services utilisés à travers les assistants vocaux et qui permettent de consulter le solde de son compte bancaire ou bien d'effectuer des virements, de consulter ses relevés, de commander les volets roulants ou les luminaires de son foyer, de suivre sa consommation énergétique, etc. La multiplication des échanges de données et des informations stockées sur différents comptes liés aux individus pose des questions pour la protection de la vie privée et des données.

1) Des données intimes, potentiellement sensibles

Comme exposé dans le Chapitre I.1 *La spécificité de la voix*, la voix peut contenir de nombreuses informations relatives aux paroles prononcées par l'individu et à son identité ou à des caractéristiques inférées telles que son état émotionnel, ses origines socio-culturelles, son ethnicité ou encore son état de santé. Dès lors, la voix dévoile des informations qui touchent à l'intime. Lorsque ces informations révèlent l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'apparte-

100 - Titiou Lecoq, *Les enfants pensent que les enceintes connectées sont vivantes et c'est un problème*, Slate.fr, juin 2018 http://www.slate.fr/story/163907 enceintes-connectees-intelligence-artificielle-alexa-google-home-siri-education-enfants
101 - CNIL, *Comment permettre à l'Homme de garder la main ? – Rapport sur les enjeux éthiques des algorithmes et de l'intelligence artificielle*, décembre 2017, https://www.cnil.fr/fr/comment-permettre-lhomme-de-garder-la-main-rapport-sur-les-enjeux-ethiques-des-algorithmes-et-de
102 - Koenecke et al., *Racial disparities in automated speech recognition, Proceedings of the National Academy of Sciences*, avril 2020, https://www.pnas.org/content/117/14/7684
103 - Antonio A. Casilli, *En attendant les robots : Enquête sur le travail du clic,* Seuil, 2019
104 - Alex Hern, *Skype audio graded by workers in China with 'no security measures'*, The Guardian, janvier 2020, https://www.theguardian.com/technology/2020/jan/10/skype-audio-graded-by-workers-in-china-with-no-security-measures

nance syndicale ou concernent l'état de santé ou la vie / l'orientation sexuelle, elles rentrent dans la catégorie des données sensibles, spécialement protégées par les lois sur la protection des données (article 9 du RGPD). La voix contient des marqueurs spécifiques à une personne, combinaisons de facteurs physiologiques et comportementaux. C'est ce qui en fait un attribut biométrique à part entière, qui peut être utilisé pour pour l'identifier. Les assistants vocaux peuvent proposer de créer des profils liés à la voix d'un individu, afin de pouvoir affilier plusieurs comptes à un même assistant vocal (par exemple les différents membres d'un foyer). En pratique, il s'agit ainsi d'identifier le locuteur parmi les profils enrôlés simultanément à la prononciation du mot-clé utilisé pour l'activation de l'assistant. Les dispositifs peuvent ainsi associer plusieurs personnes et ne pas les confondre lorsque l'un demande d'accéder à des informations qui lui sont personnelles (courriels, agenda, etc.). Toutefois, la reconnaissance de la voix est soumise à un certain nombre de contraintes : qualité du microphone, bruit ambiant, état physique de la personne, etc. qui peuvent faire obstacle à cette identification. Enfin, la circulation des données biométriques, ou permettant la constitution de modèles biométriques, et les conditions de stockage de celles-ci – potentiellement réalisé de façon distante – soulèvent des interrogations quant à leur possible récupération à des fins de piratage. Si des espaces clandestins, comme le *dark web* regorgent déjà de bases de données contenant des adresses emails, mots de passe et autres informations[105], une utilisation intense des interfaces vocales pourraient faire de ces caractéristiques biométriques un bien très échangé sur les marchés noirs des données dans le futur.

2) Des dispositifs enregistreurs de plus en plus présents dans les espaces partagés

a) Du foyer aux lieux collectifs et professionnels

Les assistants vocaux, initialement personnels et associés à un smartphone, se sont émancipés de cette notion individuelle pour se déployer dans de nouveaux espaces. D'abord présents dans le foyer avec les enceintes connectées (salon, chambre à coucher, cuisine, etc.), ils se retrouvent désormais dans des lieux de passage. Par exemple, ils sont de plus en plus régulièrement installés dans des hôtels, pour proposer de nouveaux services aux clients, ou dans des lieux professionnels, comme les cabinets médicaux, par exemple afin de faciliter la prise de rendez-vous. Ce déploiement des assistants vocaux dans ces espaces de passage questionne la confidentialité des échanges qui sont tenus, mais également le secret professionnel (notamment dans les cas liés au secret médical, dans le cadre de la relation à un avocat, pour la préservation du secret des sources, etc.). À titre d'exemple, la crainte de fuite de données ou de piratage de ces dispositifs d'écoute a conduit l'un des plus grands cabinets d'avocats d'Irlande à les interdire au sein de l'entreprise, notamment pour les salariés effectuant du télétravail et en possédant un dans leur domicile[106]. Enfin, la dissémination de ces outils dans des lieux professionnels pose également la question de la surveillance des employés. La manière dont ils sont utilisés pourrait présenter un aspect de surveillance permanent et constant[107].

b) Quelle information des personnes dans le rayon d'action de l'assistant ?

Les assistants vocaux ayant un rayon d'action de plusieurs mètres (environ cinq mètres en intérieur), l'information des personnes situées à proximité est un écueil récurrent. Alors qu'elles n'ont pas configuré le dispositif elles-mêmes, les voix de ces dernières peuvent être enregistrées, que ce soit à travers l'interpellation d'un assistant vocal dont elles ne sont pas propriétaires ou par la captation non voulue d'une conversation tenue à proximité d'un de ces dispositifs d'enregistrement. Pourtant, l'accès à l'information et l'exercice des droits doivent également être garantis pour ces personnes et plusieurs questions se posent. Comment gérer le fait que des données de tiers puissent être captées sans qu'ils n'en soient nécessairement informés ? Comment sécuriser les informations de l'utilisateur principal si des tiers sont susceptibles d'interagir avec l'objet ? Paradoxalement, l'une des solutions peut consister à utiliser encore davantage de données pour sécuriser et personnaliser certaines interactions avec l'assistant, par exemple en recourant à la reconnaissance du locuteur.

c) Des dispositifs qui brouillent la distinction entre espaces privés et publics

L'utilisation de dispositifs se reposant sur le support de la voix questionne la séparation entre l'espace privé et l'espace public et redéfinit la notion d'intimité. Parler ou recevoir des éléments audio dans un espace partagé peut avoir des conséquences diverses pour les individus : exposition accrue de sa vie privée auprès de tiers, risque de fausses manipulations et de déclenchements intempestifs, possibilités d'activations intempestives par simple prononciation du mot-clé, etc. Ces nouvelles formes d'interactions dans l'espace social modifient de fait le rapport des utilisateurs à leur vie privée.

3) Des dispositifs qui font intervenir de nombreux acteurs

Le fonctionnement des assistants vocaux implique l'intervention d'un certain nombre d'acteurs et d'intermé-

105 - Lily Hay Newman, *1.2 Billion Records Found Exposed Online in a Single Serve*, Wired, novembre 2019, https://www.wired.com/story/billion-records-exposed-online/
106 - Aaron Rogan, *Law professionals banned from working at home near Alexa devices*, BusinessPost, février 2020, https://www.businesspost.ie/legal/law-professionals-banned-from-working-at-home-near-alexa-devices-3a681a75
107 - CNIL, *La vidéosurveillance – vidéoprotection au travail*, https://www.cnil.fr/fr/la-videosurveillance-videoprotection-au-travail

diaires tout au long de la chaîne d'exécution. Entre les concepteurs d'assistants vocaux et les utilisateurs, il faut également passer par les intégrateurs, les développeurs d'applications tierces et éventuellement « les déployeurs » de ces assistants (voir Chapitre I.2 *Assistant vocal, qui es-tu ?*). La question des responsabilités engagées par chaque acteur se pose donc. Qui est responsable du traitement des données ? Comment sont organisées les relations entre le concepteur de l'assistant et les développeurs d'application ? Comme illustré dans l'infographie de la page 14, les schémas de circulation des données sont multiples et varient en fonction des usages et des choix de conception. Il convient donc de réaliser des analyses aux cas par cas afin de préciser les différents rôles, modalités d'intervention et capacités d'action de chaque acteur.

4) Les écoutes par des humains à des fins d'amélioration du produit

À la suite de nombreuses scandales publicisés à l'été 2019, tous les grands acteurs du marché des assistants vocaux (Amazon, Google, Microsoft, Facebook, et Apple) ont révélés que les enregistrements audio réalisés par leurs assistants étaient pour une partie d'entre eux réécoutés par des individus, soit directement employés par ces sociétés, soit agissant comme sous-traitants, afin de catégoriser les énonciations, de perfectionner la qualité de la détection du mot-clé, d'améliorer les performances des systèmes de transcription et d'interprétation de la parole, etc. En effet, si la pratique de contrôle et d'annotation humaine est indispensable pour les systèmes d'apprentissage automatique (*machine learning*), ces scandales soulignent l'absence d'information claire des personnes concernées quant à cette réécoute des interactions (réelles ou supposées) passées avec un assistant. En réponse, l'autorité de protection des données d'Hambourg a enjoint la société Google de suspendre de façon provisoire (pendant une période de trois mois) les activités liées à l'écoute et à la transcription manuelle des paroles collectées par son assistant[108]. Par la suite tous les grands acteurs du marché des assistants vocaux ont annoncé avoir mis fin à ces pratiques ou mis en place un système d'opt-out pour les utilisateurs[109] [110]. Ces écoutes à des fins d'amélioration du dispositif ont parallèlement soulevé d'autres interrogations. Certaines personnes exerçant ces métiers basés sur l'écoute des échanges ont indiqué avoir entendu des enregistrements qualifiés de dérangeants, voire relevant d'activités illégales, allant jusqu'à l'agression[111] [112]. Enfin, la saisie par des tribunaux d'enregistrements réalisés par des assistants vocaux lors d'enquêtes criminelles pose également des questions sur les implications juridiques que l'usage d'assistants vocaux pourrait entraîner[113].

5) Faux positifs et personnes extérieures : la question du consentement et de la maîtrise de ses données

Le RGPD a pour ambition de redonner la souveraineté de ses données et les capacités d'exercice de ses droits à l'individu. Dans le cas des assistants vocaux, le fonctionnement même du dispositif nécessite un contrôle à posteriori. L'individu n'a accès à son activité et aux enregistrements (dans le cas où ceux-ci sont conservés) qu'une fois que ceux-ci ont été traités et, selon les implémentations, envoyés sur des serveurs de traitement distants. Il est alors possible de réécouter ou de supprimer les enregistrements, y compris certains faux positifs, c'est-à-dire des déclenchements intempestifs de l'assistant suite à une fausse détection du mot-clé. Le simple fait de pouvoir savoir ce qui a été transmis aux serveurs seulement en allant consulter son compte interroge, qui plus est quand ces données n'avaient pas vocation à être transmises. La perte de contrôle du compte ou son piratage peut donc entraîner un accès non seulement à une activité consentie (lors d'une utilisation volontaire du dispositif), mais également à une partie plus personnelle, qui peut être plus intime et n'est pas forcément connue du propriétaire de l'assistant. À ceci peut s'ajouter l'intervention de personne n'ayant pas connaissance du dispositif d'enregistrement, et dont les paroles sont captées malgré eux. Elles peuvent se retrouver conservées à leur insu et rester également accessibles au propriétaire.

6) Privacy by design et initiatives vertueuses

Certains concepteurs d'assistants vocaux font le choix de développer des produits respectueux de la vie privée et de la protection des données personnelles à l'instar de la startup Snips (racheté par Sonos en 2019) ou encore Mycroft, qui a développé une solution d'assistant vocal en *open source*[114].

108 - The Hamburg Commissioner for Data Protection and Freedom of Information, *Speech assistance systems put to the test - Data protection authority opens administrative proceedings against Google*, août 2019, https://datenschutz-hamburg.de/assets/pdf/2019-08-01_press-release-Google_Assistant.pdf
109 - Guillaume Périssat, *Assistants vocaux : Apple et Google suspendent les écoutes tandis qu'Amazon propose de les désactiver*, L'informaticien, août 2019, https://www.linformaticien.com/actualites/id/52597/assistants-vocaux-apple-et-google-suspendent-les-ecoutes-tandis-qu-amazon-propose-de-les-desactiver.aspx
110 - Lidia Davis, *How to opt out of Google Home's tracking features*, Reviews.com, mars 2020, https://www.reviews.com/home/smart-home/opt-out-of-google-homes-tracking-features/
111 - Matt Day, Giles Turner, and Natalia Drozdiak, *Amazon Workers Are Listening to What You Tell Alexa*, Bloomberg, avril 2019, https://www.bloomberg.com/news/articles/2019-04-10/is-anyone-listening-to-you-on-alexa-a-global-team-reviews-audio
112 - Anouk Helft, *Apple admet écouter certaines conversations privées via Siri*, Les Echos, juillet 2019 https://www.lesechos.fr/tech-medias/hightech/apple-admet-ecouter-certaines-conversations-privees-via-siri-1041502
113 - LePoint.fr, *L'enceinte Alexa a-t-elle été témoin d'un meurtre ?*, Le Point, novembre 2019 https://www.lepoint.fr/monde/l-enceinte-alexa-a-t-elle-ete-temoin-d-un-meurtre-07-11-2019-2345984_24.php
114 - https://mycroft.ai/

Dans son article 25, le RGPD rappelle l'importance de concevoir des systèmes par construction respectueux de la vie privée des utilisateurs. Adapté au cas des assistants vocaux, il est possible de construire une approche du cycle de la donnée plus soucieuse des droits des individus. En effet, il semble pertinent de catégoriser les types d'utilisation de l'assistant et d'y associer des modèles permettant une minimisation des données transférées – et, dans une certaine mesure, de sobriété numérique. D'abord, pour les usages internes (prendre des notes, mettre un minuteur, etc.) ou domotiques (allumer la lumière, baisser les volets roulants, etc.), un traitement local de l'information peut présenter moins de risques pour la vie privée des utilisateurs qu'un traitement se reposant sur l'utilisation de serveurs distants pour un résultat identique. Un tel choix de conception suppose d'embarquer plus « d'intelligence » (capacités calculatoires) dans les dispositifs. Dans une telle optique, les relations et échanges avec des éléments extérieurs à l'équipement embarquant l'assistant vocal seraient réduits à deux grands usages qui relèvent 1) des actions nécessitant d'accéder à des bases de connaissance distantes (accéder à une information en ligne, prendre un rendez-vous, passer un appel, écouter de la musique sur un site de streaming, etc.) et 2) de ce qui concerne la remontée d'information à des fins d'amélioration du produit ou de statistiques d'usage.

Impact environnemental, modèle économique, position dominante et concurrence : d'autres enjeux pour les assistants vocaux

Comme de très nombreux dispositifs, l'impact environnemental des assistants vocaux pose question, notamment par la consommation énergétique qu'ils induisent. Que ce soit en termes de production (par la fabrication massive de ces nouveaux dispositifs), de transit de données (par un fonctionnement se basant sur serveur distant) ou de capacité de calcul nécessaire au traitement de la voix, les assistants vocaux sont fortement consommateurs d'énergie. Ainsi, comme mis en lumière par le projet ANR DAPCODS/IOTics (voir encadré page 45), le trafic généré par une enceinte se compte en centaines de kilo-octets, soit deux ordres de grandeur de plus que pour des objets connectés ne disposant pas de capacités de traitement automatique de la parole. Dès lors, demander à son assistant vocal d'allumer une lampe n'a pas le même impact que d'utiliser un interrupteur – et cette différence n'est pas forcément perceptible pour l'utilisateur. Qui plus est, les modèles de langage naturel utilisés dans les assistants vocaux se sont considérablement améliorés au prix d'une augmentation considérable du coût énergétique. À titre d'exemple, l'entrainement par Google du model BERT (voir encadré page 19) a nécessité l'apprentissage de quelques 340 millions de paramètres pour un coût en électricité équivalent à la consommation d'un ménage américain pendant 50 jours. Aujourd'hui, certains modèles de langage tels que GPT-2 d'OpenAI ou MegatronLM de Nvidia comptent des milliards de paramètres[115] ! Ces exemples illustrent bien ce coût du numérique qu'il est désormais de prendre en compte, y compris dans le champ de la protection des données personnelles. Dans une certaine mesure, le principe de minimisation des données fait écho au concept de sobriété numérique et vise également à créer des systèmes optimisés dans l'utilisation des données. La CNIL s'est d'ailleurs lancée dans une réflexion sur l'urgence climatique et les enjeux de régulation des nouvelles technologies dans une démarche coordonnée avec les autres autorités publiques et administratives indépendantes[116].

Par ailleurs, l'utilisation et la diffusion toujours plus grande des assistants vocaux dans notre quotidien, qu'il s'agisse d'espaces publics comme privés, soulèvent également des enjeux économiques et de concurrence. Il faudrait ainsi s'interroger sur la possibilité d'alternatives aux offres déjà existantes des géants du numérique, qu'ils soient américains ou chinois. La prédominance de ces acteurs s'illustre dans l'actualité : l'enceinte Djingo, création commune des opérateurs Orange et Deutsche Telekom, embarque par exemple l'assistant d'Amazon, Alexa.

D'un point de vue économique, la question de la rémunération des différents acteurs de la chaîne de production se pose également. Quel est le modèle économique sous-jacent ? Pourrait-on avoir dans les années à venir un modèle payant pour améliorer le référencement de son application ? Le rapport CSA-Hadopi pose la question de la répartition des profits entre éditeurs d'applications, de contenus et les services de publicité gérés par les constructeurs. Les bénéfices sont-ils équitablement répartis entre ces acteurs ? Enfin, la massification des bases de données des principaux acteurs du marché génère des réactions et inquiétudes. Quid de la concurrence et de potentiels abus de position dominante ?

115 - Karen Hao, *Tiny AI models could supercharge autocorrect and voice assistants on your phone*, MIT Technology Review, octobre 2019, https://www.technologyreview.com/2019/10/04/132755/tiny-ai-could-supercharge-autocorrect-voice-assistants-on-your-phone/

116 - Autorité de la concurrence, AMF, Arcep, ART, CNIL, CRE, CSA, HADOPI, *Accords de Paris et Urgence Climatique : Enjeux de Régulation*, https://www.arcep.fr/fileadmin/user_upload/publications/cooperation-AAI/publication_AAI-API_Accord_de_Paris_052020.pdf

Entretien avec... EMMANUEL VINCENT

Emmanuel Vincent *est Directeur de Recherche au sein de l'équipe Multispeech (Université de Lorraine, CNRS, Inria)*[117]*. Ses recherches portent notamment sur la commande vocale mains-libres et l'analyse des sons ambiants. Il développe des technologies d'intelligence artificielle peu gourmandes en données et respectueuses de la vie privée. Il coordonne le projet COMPRISE et est l'un des organisateurs du défi VoicePrivacy.*

Les assistants vocaux sont-ils condamnés à exposer la vie privée des utilisateurs ? Des implémentations protectrices de la vie privée sont-elles envisageables ? À l'occasion de la rédaction de ce Livre Blanc et dans le cadre du partenariat CNIL-Inria, le LINC s'est entretenu, avec Emmanuel Vincent, chercheur Inria et dont les travaux portent sur le développement de nouvelles interfaces vocales satisfaisants dès la conception les impératifs de protection des données.

De très nombreuses études annoncent une adoption massive des assistants vocaux dans les années à venir. Quels seront, selon vous, les enjeux pour leurs utilisateurs lorsque tous nos équipements en seront équipés ?

En permettant aux utilisateurs d'exprimer des demandes complexes, les assistants vocaux répondent au besoin d'interaction efficace avec les contenus Internet, les objets et les services du quotidien. Les entreprises de technologies vocales vont élargir les langues prises en charge et combiner la commande vocale à l'analyse d'autres aspects de la voix (âge, émotions, préférences, etc.) afin de mieux caractériser l'utilisateur et ses désirs. Les entreprises de toutes sortes vont à leur tour intégrer ces technologies dans un nombre croissant de produits.

Cela soulève de nombreux enjeux pour les citoyens, les entreprises utilisatrices et les pouvoirs publics. Par exemple, la prise en charge d'une langue a un coût qui n'est pas toujours commercialement rentable. Il est essentiel pour la diversité culturelle et l'égalité des chances de soutenir les initiatives de logiciels libres et de données ouvertes, afin que ces technologies deviennent accessibles à tous les citoyens, quelle que soit leur langue, leur dialecte ou leur accent. Il est aussi essentiel que les réponses apportées par les assistants soient équitables et explicables : en réponse à une question sur un produit, pourquoi mettre en avant les sites web de certaines marques plutôt que d'autres ?

Les usages doivent être contrôlés : une technologie comme l'analyse des émotions peut être à la fois bénéfique pour fluidifier l'interaction à un instant donné et éthiquement répréhensible si les émotions détectées sont conservées à des fins de profilage commercial. Même lorsque l'usage est acceptable, la collecte de données vocales pose des questions de sécurité et de confidentialité.

Il est donc nécessaire d'anticiper les fonctionnalités et les usages futurs afin de construire le cadre légal approprié et de permettre aux citoyens de devenir des utilisateurs avertis.

Le Règlement général sur la protection des données (RGPD) prône une approche de protection de la vie privée dès la conception (*privacy by design*). Comment un tel concept se traduit-il concrètement dans le cas des assistants vocaux ?

Selon le RGPD, la voix est une donnée personnelle. Elle véhicule en effet quatre types d'information de nature personnelle : les mots prononcés, les caractéristiques biométriques de la personne qui les a prononcés (identité,

117 - https://team.inria.fr/multispeech

âge, genre…), la façon dont elle les a prononcés (émotions et pathologies se traduisant dans la voix) et l'environnement dans lequel elle les a prononcés (voix et bruits ambiants). Le RGPD va plus loin en catégorisant comme des informations de nature sensible les caractéristiques biométriques et les mots trahissant l'orientation sexuelle ou les opinions religieuses par exemple.

Concrètement, les assistants vocaux demandent l'autorisation expresse aux utilisateurs d'utiliser leur voix pour certains usages prédéfinis et leur offrent la possibilité d'accéder aux données enregistrées et de demander leur suppression. Cela est conforme à la loi, mais ne permet pas aux utilisateurs de contrôler finement les usages qui sont faits de leurs données, dans la mesure où les usages prédéfinis ne sont souvent pas aussi spécifiques que les utilisateurs avertis pourraient le souhaiter.

Faisant notamment suite aux travaux sur la théorie de l'information de Claude Shannon, la recherche dans le domaine du traitement automatique de la parole remonte aux années 1960 environ. Toutefois, il semble que le fait d'allier celle-ci à des techniques de protection de la vie privée soit encore très récent. Quelle en est la raison ?

Les technologies vocales fonctionnent par apprentissage automatique à partir d'enregistrements de voix retranscrits sous forme textuelle. Pendant longtemps, ces données étaient acquises auprès de sujets volontaires et les systèmes ne fonctionnaient de façon suffisamment fiable que pour la reconnaissance de chiffres ou de mots-clés, qui est peu critique pour la vie privée.

Le boom des assistants vocaux est dû à la conjonction de trois facteurs : l'émergence de méthodes d'apprentissage plus puissantes, l'augmentation de la capacité de calcul et l'explosion de la quantité de données. Certaines entreprises conservent toutes les commandes vocales envoyées à leur assistant dans divers cas d'usage et s'en servent notamment pour l'apprentissage. Cette augmentation de la quantité et de la diversité des données de chaque utilisateur associée à l'augmentation de la capacité à en extraire des informations accroît les risques pour la vie privée, que ce soit dans le cadre d'un usage légal ou illégal (cyberattaque) : profilage, accès à des informations sensibles, usurpation d'identité, espionnage industriel, etc. Le profilage est une pratique courante, qui pourrait se voir renforcée par le recoupement d'informations issues de multiples cas d'usage. Les autres risques peuvent sembler exagérés aujourd'hui, mais constituent une menace probable à un horizon de quelques années.

Pour limiter ces risques, d'autres entreprises font le choix de ne pas conserver les commandes vocales effectuées et d'utiliser des données d'apprentissage acquises auprès de sujets volontaires, au risque que leurs produits soient moins efficaces.

Vous menez des recherches sur le sujet. Pouvez-vous nous présenter comment vous en êtes arrivé à travailler sur ces objets et les défis que vous souhaitez relever ?

Mon intérêt découle du constat que, pour atteindre les bienfaits économiques et sociétaux attendus de l'intelligence artificielle et des assistants vocaux notamment, nous devons développer des outils d'apprentissage automatique efficaces capables de tirer le meilleur de données personnelles massives tout en garantissant la préservation de la vie privée, de l'équité et des autres valeurs auxquelles sont attachées nos concitoyens. Le déclic est venu du contact avec l'équipe Magnet (Université de Lille, CNRS, Inria), qui conçoit de tels outils et apporte des garanties formelles de confidentialité, et avec des entreprises européennes, qui ont exprimé leur intérêt.

Depuis fin 2018, dans le cadre du projet COMPRISE financé par le programme Horizon 2020 de l'Union européenne, nous concevons un assistant vocal *open source* et une plateforme d'apprentissage fondés sur le principe de protection de la vie privée par défaut[118]. Pour cela, avant d'envoyer les données de l'utilisateur vers la plateforme d'apprentissage, nous transformons la voix et remplaçons certains mots afin que l'utilisateur ne soit plus identifiable. Nos premiers essais nous ont donné du fil à retordre car les outils de biométrie moderne sont extrêmement puissants pour ré-identifier l'utilisateur, même après transformation. Nos outils ne garantissent pas une anonymisation parfaite, mais fournissent un niveau de protection très supérieur à l'existant.

Mon équipe coordonne aussi le projet DEEP-PRIVACY financé par l'Agence nationale de la recherche, qui adopte une approche alternative d'apprentissage décentralisé. Dans cette approche, les données personnelles ne quittent pas le terminal de l'utilisateur, ce qui fournit une protection accrue mais a l'inconvénient de ne plus permettre leur retranscription manuelle. Pour susciter d'autres initiatives de ce genre, nous avons créé le défi VoicePrivacy dont les résultats seront présentés très prochainement[119].

118 - COMPRISE, *Cost-effective Multilingual, Privacy-driven voice-enabled Services*, (2019 – 2022), https://www.compriseh2020.eu/
119 - https://www.voiceprivacychallenge.org/

ZOOM SUR...

Assistants vocaux et recherche en protection de la vie privée

Ce n'est que depuis une dizaine d'années environ que les recherches en traitement de la parole et en vie privée ont commencé à être associées. Elles se sont tout d'abord focalisées sur l'utilisation de techniques de cryptographie appliquées au traitement du signal telles que le calcul multipartite sécurisé, le transfert inconscient ou le chiffrement homomorphe[120] [121]. Plus récemment, parallèlement à l'adoption croissante des assistants vocaux, on a assisté à la multiplication de travaux de recherche et de publications. Si les aspects relatifs à la sécurité de ces dispositifs équipés d'assistants vocaux ont déjà largement été évoqués précédemment (voir page 35), de nombreux projets s'inscrivant au carrefour du traitement de la parole et de la protection de la vie privée ont également été financés.

À titre d'exemple, le projet PAMELA financé par l'Agence nationale de la recherche (ANR) vise à développer des méthodes d'apprentissage automatique utilisant des modèles personnalisés locaux, de façon décentralisée et coopérative au sein de réseaux où sont distribués les données et systèmes apprenants[122]. Un des cas d'usages étudié est notamment celui d'un assistant vocal respectueux de la vie privée (la société SNIPS était membre du consortium). Comme évoqué par Emmanuel Vincent dans son interview (voir page 43), les projets ANR DEEP-PRIVACY et H2020 COMPRISE s'inscrivent dans cette lignée, en proposant également de s'intéresser aux techniques d'apprentissage fédéré et de confidentialité différentielle[123] [124]. Des approches « adversaires » sont déployées pour modifier le signal de parole afin qu'il ne soit plus porteur des caractéristiques biométriques de l'utilisateur sans pour autant que la qualité de la transcription automatique ne soit dégradée (*speech to text*). Le projet franco-japonais VoicePersonae vise quant à lui à servir de « hub » à de nombreux sujets liés à l'identité vocale tels que la synthèse de la parole, la reconnaissance du locuteur, la détection des attaques de présentation (*spoofing*), la criminalistique des médias, l'anonymisation de la parole, etc.

Afin de fédérer la communauté de recherche autour de ces sujets et de mesurer les progrès dans un esprit d'émulation, des défis scientifiques ont également été lancés. Il s'agit ainsi de proposer à différentes équipes de recherche de se mesurer les unes aux autres sur la réalisation d'une tâche préalablement définie. Ainsi, le défi ASVSpoof promeut la production de contre-mesures pour lutter contre l'usurpation d'identité dans le cadre de l'authentification vocale[125]. Le défi VoicePrivacy vise quant à lui à développer des solutions d'anonymisation pour supprimer les informations personnelles contenues dans les signaux vocaux[126].
De façon connexe, le projet ANR DAPCODS/IOTics s'intéresse aux objets connectés de la maison intelligente sous l'angle de la protection des données. Ce travail porte en particulier sur les moyens de contrôle de la maison dont dispose un utilisateur, au premier rang desquels figurent l'assistant vocal de son smartphone et enceinte. En effet, ce sont généralement ces moyens de contrôle qui déterminent la nature des données échangées (qui vont parfois au-delà des données collectées par l'objet), leurs volumes et les acteurs impliqués, ce qui pose également des questions de souveraineté[127].

Enfin, la Chaire HUMAAINE (*human-machine affective interaction & ethics*) étudie les interactions et relations humain-machine afin d'auditer et de mesurer l'influence potentielle des systèmes affectifs[128]. Les travaux portent sur la détection des émotions sociales dans la voix humaine et, en utilisant les contributions de l'économie comportementale mises en évidence par le prix Nobel Richard Thaler, sur l'étude des manipulations dans le langage parlé et audio qui sont destinées à induire des changements dans le comportement de l'interlocuteur humain. L'objectif final de ces travaux scientifiques est de permettre la conception de systèmes éthiques *by design*.

120 - R. (Inald) L. Lagendijk, Zekeriya Erkin et Mauro Barni, *Encrypted Signal Processing for Privacy Protection*, IEEE Signal Processing Magazine, janvier 2013.
121 - Manas Pathak, *Privacy-Preserving Machine Learning for Speech Processing*, Springer, 2013.
122 - PAMELA, *Personalized and decentralized machine learning under constraints*, (2016 – 2020), https://project.inria.fr/pamela/
123 - DEEP-PRIVACY, *Apprentissage distribué, personnalisé, préservant la privacité pour le traitement de la parole*, (2018 – 2022), https://anr.fr/Projet-ANR-18-CE23-0018
124 - COMPRISE, *Cost-effective multilingual, privacy-driven voice-enabled services,* (2019 – 2022), https://www.compriseh2020.eu/
125 - https://www.asvspoof.org/
126 - https://www.voiceprivacychallenge.org/
127 - DAPCODS/IOTics, *Data protection of connected devices and smartphones*, (2017-2020), https://project.inria.fr/iotics/fr/
128 - http://humaaine-chaireia.fr/

CAS D'USAGES : LE RGPD EN PRATIQUE

Dans ce chapitre, il est proposé d'étudier la manière dont le RGPD peut s'appliquer dans plusieurs contextes d'usage des assistants vocaux, qu'il s'agisse d'un assistant personnel (sur smartphone par exemple) comme de tout dispositif pouvant être intégré dans un cadre plus collectif (composant par exemple le foyer : téléviseur connecté, réfrigérateur connecté, enceinte de salon, etc.).

L'objectif est de fournir aux utilisateurs particuliers ou professionnels, mais également aux concepteurs d'assistants et aux développeurs d'application destinés à ces assistants, des exemples à suivre leur permettant de se mettre en conformité avec les différents aspects de la réglementation.

Les cas d'usages présentés ci-après sont des cas fictifs inspirés de ce qui est actuellement proposé sur le marché des assistants vocaux. Ils ne constituent pas une analyse exhaustive et complète de la manière dont le RGPD doit être appliqué dans le contexte des assistants vocaux, mais donnent des pistes d'analyse et de réflexion. Ils se basent sur des postulats qui ne reflètent pas nécessairement les modalités de fonctionnement de tous les assistants vocaux. Ainsi, celles-ci reposent sur l'hypothèse que les principales fonctions de traitement de la parole et des actions se réalisent en ligne, sur des systèmes maîtrisés par les fournisseurs de solution, reflétant ainsi le fonctionnement de la plupart des assistants vocaux disponibles actuellement pour le grand public – mais d'autres modèles de traitement sont possibles.

Trois cas sont présentés afin de couvrir différentes problématiques :

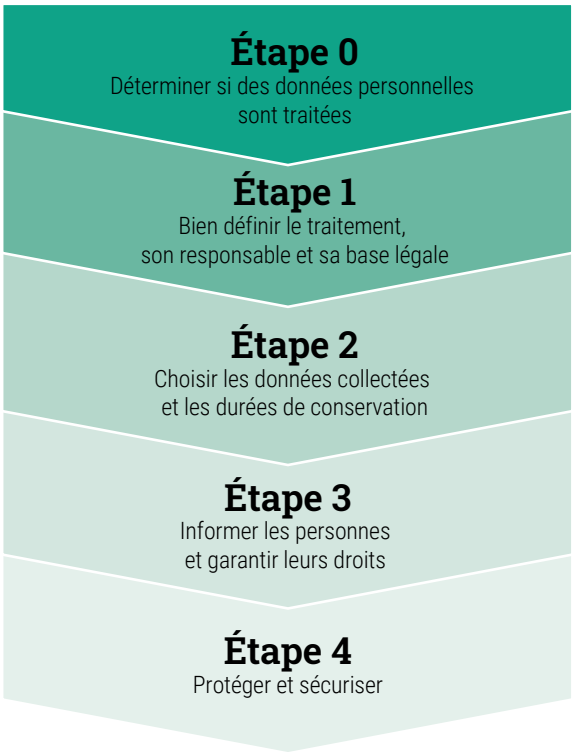
CAS n°1 :
Utiliser les fonctions de base de son assistant vocal

CAS N°2 :
Utiliser une application bancaire via un assistant vocal

CAS N°3 :
Réutiliser les données collectées par l'assistant vocal à des fins d'amélioration du service

En effet, une distinction doit être opérée entre l'interrogation de l'assistant sur une question d'ordre général pour laquelle il consulte des ressources publiquement accessibles sur Internet, l'utilisation d'une application ou encore la gestion d'équipements connectés par l'intermédiaire de l'assistant. Le premier cas ne nécessite pas, outre la création d'un compte utilisateur, l'installation et/ou l'utilisation d'une application spécifique. Un seul acteur intervient : le concepteur de l'assistant. À l'inverse dans le deuxième et troisième cas, plusieurs acteurs peuvent intervenir, à savoir le concepteur de l'assistant, le développeur de l'application pour le produit acheté et/ou le fournisseur de l'équipement connecté (par exemple, une ampoule connectée).

Ainsi, la mise en conformité des traitements de l'assistant vocal peut être analysée en suivant quatre étapes principales (l'étape 0 étant un préalable) :



Concernant l'étape 0, la première question à se poser est de déterminer si les assistants vocaux traitent des données personnelles. En effet, on pourrait se demander si la collecte de simples phrases relève effectivement de la protection des données personnelles. La réponse est directement traitée dans cette introduction et est évidemment positive. Comme décrit dans le Chapitre I *Assistants vocaux, de quoi parle-t-on ?*, la voix est incontestablement une donnée personnelle. De plus, les questions posées ou même la simple activité enregistrée par l'assistant vocal pendant sa phase de veille peuvent révéler des informations personnelles sur la personne comme sa localisation, ses préférences, ses heures de présence, etc. Enfin, la transmission des données enregistrées par l'assistant est accompagnée de données personnelles, soit indirectement identifiantes (numéro d'utilisateur ou d'équipement, adresse IP, pseudo, etc.), soit directement identifiantes (données du compte, adresse courriel, etc.).

Les notions clés du RGPD

Collecter et traiter des données personnelles implique de suivre les principes cardinaux de la protection des données et notamment d'informer les personnes sur ce qui est fait de leurs données et de respecter leurs droits. En tant que responsable d'un traitement de données, ou en tant que sous-traitant, il est nécessaire de prendre des mesures pour garantir une utilisation de ces données respectueuse de la vie privée des personnes concernées. Les principes suivants s'appliquent à tout traitement de données personnelles.

Finalité & Statut des Acteurs : c'est l'objectif principal de l'utilisation de données personnelles. En effet, il est obligatoire de définir préalablement le but précis dans lequel les données sont recueillies ou traitées. La finalité est à respecter tout au long du cycle de vie des données. On qualifie de responsable de traitement la personne morale (entreprise, commune, etc.) ou physique qui détermine les finalités et les moyens du traitement.

Base Légale : il est nécessaire d'associer à chaque finalité un fondement juridique : contrat, consentement, intérêt légitime, obligation légale, mission d'intérêt public ou sauvegarde des intérêts vitaux. Les détails de choix et d'application de chacune d'entre elles sont à retrouver sur le site de la CNIL[129].

Exactitude, Proportionnalité & Minimisation des Données : seules les informations adéquates, pertinentes et strictement nécessaires au regard des objectifs préalablement fixés sont utilisées. Les données doivent être exactes et tenues à jour. Les données inexactes doivent être rectifiées ou effacées.

Limitation de la Durée de Conservation : la durée de conservation doit être fixée en amont, et respecter un principe de proportionnalité et d'équilibre qui dépend des finalités poursuivies. Elle ne peut être illimitée et ne doit durer que le temps nécessaire pour atteindre l'objectif (la finalité) préalablement fixé et porté à la connaissance des personnes concernées, en conformité avec les autres obligations légales s'appliquant éventuellement (cadre légale du secteur bancaire, assurance, santé, etc.).

Sécurité : il est obligatoire de prendre les mesures appropriées de sécurité, informatique mais aussi physique, pour garantir l'intégrité, la disponibilité et la confidentialité des données. Elles doivent être adaptées en fonction de la sensibilité des données et des risques qui pèsent sur les personnes en cas d'incident.

Information & Transparence : il est indispensable de fournir une information concise, transparente, compréhensible et aisément accessible (voir l'encadré Données & Design page 69). La transparence constitue le socle du contrat de confiance qui lie les organismes avec les personnes dont ils traitent les données. C'est également un des droits fondamentaux des personnes (voir plus bas).

Maîtrise des Données & Identification des Risques : le partage et la circulation des données personnelles doivent être encadrés et contractualisés, afin de leur assurer une protection à tout moment. Des mesures spécifiques peuvent s'appliquer dans les cas où de grands volumes de données ou des données sensibles sont traitées ou si le traitement de données peut avoir des conséquences particulières pour les personnes. Parmi ces mesures, on trouve l'Analyse d'Impact relative à la Protection des Données (ou AIPD), qui est obligatoire dans certains cas (la liste des critères et la liste des types d'opérations concernées sont disponibles sur le site de la CNIL[130]) – voir l'encadré sur la méthode PIA appliquée au domaine des objets connectés page 56.

129 - CNIL, Les bases légales, https://www.cnil.fr/fr/les-bases-legales
130 - CNIL, Ce qu'il faut savoir sur l'analyse d'impact relative à la protection des données, https://www.cnil.fr/fr/ce-quil-faut-savoir-sur-lanalyse-dimpact-relative-la-protection-des-donnees-aipd

Protection des Données Sensibles : ce sont des informations qui révèlent la prétendue origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique. Ces données sont soumises à une réglementation : elles ne peuvent pas être collectées et traitées sauf exceptions (notamment par le consentement explicite de l'individu – voir page 53 et le site de la CNIL pour plus d'informations[131]).

Droits des Personnes : les personnes concernées disposent de droits afin de garder la maîtrise de leurs données. Le responsable de traitement doit leur expliquer comment les exercer (auprès de qui, sous quelle forme, etc.). Lorsqu'elles exercent leurs droits, les personnes doivent obtenir une réponse avant un mois. Ces droits sont au nombre de huit :

- **Droit à l'information :** permet à l'individu d'être conscient du traitement des données personnelles le concernant et d'exercer ses autres droits.
- **Droit d'accès :** permet à l'individu de savoir lesquelles de ses données sont traitées et d'en obtenir la communication dans un format compréhensible. Il permet également de contrôler l'exactitude des données et, au besoin, de les faire rectifier ou effacer.
- **Droit de rectification :** permet de corriger ses données inexactes (par exemple, âge ou adresse erronés) ou de compléter des données (par exemple, adresse sans le numéro de l'appartement) en lien avec la finalité du traitement.
- **Droit d'opposition :** permet de s'opposer à ce que ses données soient utilisées par un organisme pour un objectif précis. Cette opposition doit être motivée, sauf en cas de prospection commerciale à laquelle il est possible de s'opposer sans motif.
- **Droit à l'effacement :** permet d'effacer ses données (à certaines conditions).
- **Droit à la portabilité :** permet de récupérer les données fournies à un organisme dans le cadre de l'usage de son service, dans un format numérique communément utilisé, pour un usage personnel ou pour les transmettre à un tiers de son choix.
- **Droit à l'intervention humaine :** chaque individu a le droit de ne pas faire l'objet d'une décision entièrement automatisée, lorsqu'elle produit des effets juridiques ou l'affecte de manière significative. Une telle décision ne peut être déclenchée que dans les conditions précisées par le RGPD et en préservant le droit de la personne concernée d'obtenir une intervention humaine, d'exprimer son point de vue et de contester la décision.
- **Droit à la limitation du traitement :** permet de compléter d'autres droits cités ci-dessus (rectification, opposition, etc.). Si l'exactitude des données utilisées est contestée par l'organisme ou que l'individu s'oppose à ce que ses données soient traitées, la loi autorise l'organisme à procéder à une vérification ou à examen de la demande pendant un certain délai. Au cours de celui-ci, l'individu a la possibilité de demander à l'organisme de geler l'utilisation de ses données.

131 - CNIL, Définition – Donnée sensible, https://www.cnil.fr/definition/donnee-sensible

CAS N°1 :
Utiliser les fonctions de base de son assistant vocal

Le premier usage connu de l'assistant vocal vise à répondre simplement et rapidement à des besoins fonctionnels récurrents. Ainsi, pour les assistants vocaux dits « généralistes » (voir Chapitre I.4 *Quelle(s) stratégie(s) pour les concepteurs d'assistants vocaux ?*), un utilisateur peut interroger son assistant sur divers sujets comme il le ferait en utilisant un moteur de recherche sur le Web. Il peut par exemple consulter la météo, se renseigner sur le trajet le plus court pour se rendre à son travail ou encore demander à allumer la radio.

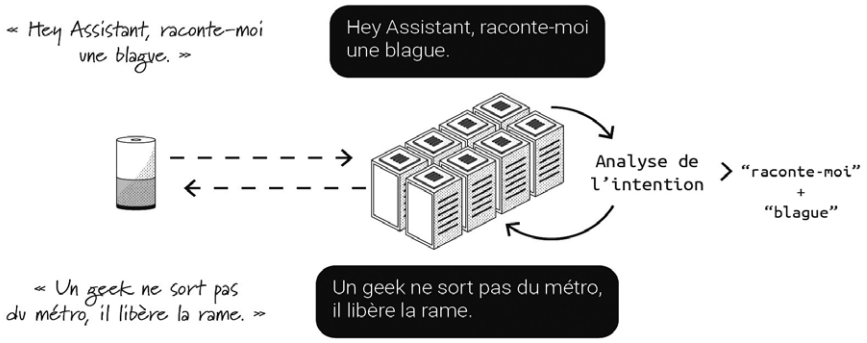


Figure 4
Mise en œuvre des fonctionnalités de base d'un assistant vocal

En pratique, comme illustré dans la Figure 4, le fonctionnement de l'assistant vocal est le suivant (plus de détails dans l'infographie page 14) :

- Captation de la requête de l'utilisateur suite à l'activation de l'assistant ;
- Transmission aux serveurs du concepteur de l'assistant pour transcription automatique de la parole, interprétation de la commande et identification d'une réponse adaptée ;
- Déclenchement à distance d'une réponse ou d'une action (ici « raconter une blague ») qui est exécutée in fine via l'équipement intégrant l'assistant.

Dans la majorité des cas, et avant de pouvoir faire fonctionner l'assistant vocal, l'utilisateur doit disposer d'un compte associé à l'appareil. Il devra créer un compte spécifique ou le concepteur de l'assistant lui permettra de lier un compte déjà existant pour utiliser directement le produit.

NOTA BENE :

il n'est pas techniquement nécessaire d'utiliser un compte utilisateur pour se servir de l'assistant vocal afin d'accéder à des informations disponibles publiquement en ligne (météo, actualités, etc.), de la même manière qu'il n'est pas techniquement nécessaire d'en créer un pour pouvoir naviguer sur le Web.

Avertissements sur l'étude de cas n°1

L'étude de ce cas « générique » (poser des questions, demander la météo, etc.) permet de poser les bases juridiques de la relation entre l'utilisateur et le concepteur de l'assistant vocal. Les grands principes déclinés (minimisation, limitation, base légale, information, etc.) demeurent pour les deux autres cas présentés par la suite : seules les spécificités relatives à chaque cas seront ensuite développées.

Étape 1 :
Bien définir le traitement, son responsable et sa base légale

Définir la finalité et le statut des acteurs

Dans ce cas d'usage, les traitements concernent à la fois les données du compte utilisateur, les commandes adressées par l'utilisateur à l'assistant (la voix et sa retranscription sous forme de texte) et les données nécessaires au traitement de la demande (préférences, localisation, date et heure, etc.).

Tel que défini dans le RGPD, le responsable de traitement est la personne morale (entreprise, commune, etc.) ou physique qui détermine les finalités et les moyens d'un traitement, c'est à dire l'objectif et la façon de le réaliser. Ici, le concepteur de l'assistant est donc le responsable de traitement dans la mesure où il en détermine les finalités (la fourniture du service d'assistance vocale) et les moyens (le traitement via l'assistant relié à un compte utilisateur).

Préciser la base légale du traitement de données

La définition de la base légale permet au concepteur de l'assistant vocal de déterminer le fondement juridique des traitements de données. Dans ce cas d'usage où l'utilisation de l'assistant est limité aux fonctionnalités fournies par le concepteur, les traitements de données décrits ci-dessus sont nécessaires pour exécuter le service demandé par l'utilisateur via l'assistant vocal. Dès lors, leur base légale pourrait être l'exécution d'un contrat auquel l'utilisateur, personne concernée, est partie (article 6(1.b) du RGPD).

Dans ce cas d'usage, la base légale liée au consentement des personnes concernées paraît relativement inadaptée. En effet, pour être valide au regard du RGPD, le consentement doit être libre, spécifique, éclairé et univoque. Dans le cas des assistants vocaux, le défaut du consentement entrainerait l'impossibilité de profiter de leurs services. Par exemple, le défaut du consentement au traitement de la voix ne permettrait pas dans ce cas d'effectuer sa transcription automatique. Cette conséquence négative pourrait peser dans la décision de la personne concernée de donner ou non son consentement, qui ne serait alors pas libre.

ZOOM SUR...

Les données nécessaires à l'exécution du contrat

Le contrat ne peut valablement fonder un traitement de données que si ce dernier est objectivement nécessaire à l'exécution du contrat. Il ne suffit pas que le traitement de données soit mentionné dans des clauses contractuelles ou dans des conditions générales d'utilisation. En d'autres termes, le traitement de données doit uniquement permettre à l'organisme de fournir le produit ou le service souhaité par l'utilisateur, et ne doit pas viser un autre objectif permettant, par exemple, la poursuite d'intérêts distincts ou exclusifs du concepteur de l'assistant vocal (pour plus d'informations, voir la page dédiée sur le site de la CNIL[132]). Si les données sont utilisées pour d'autres finalités, la base légale pour ces finalités devra être différente.

> ***« Il n'est pas toujours techniquement nécessaire d'utiliser un compte utilisateur pour se servir d'un assistant vocal. »***

132 - CNIL, Le contrat : dans quels cas fonder un traitement sur cette base légale ?, 2020, https://www.cnil.fr/fr/le-contrat-dans-quels-cas-fonder-un-traitement-sur-cette-base-legale

ZOOM SUR...

Le profilage publicitaire et la directive ePrivacy

Comme vu dans le Chapitre I.4 *Quelle(s) stratégie(s) pour les concepteurs d'assistants vocaux ?*, il est possible que certains concepteurs d'assistants vocaux souhaitent utiliser l'historique des demandes adressées par l'utilisateur à l'assistant pour créer et alimenter un profil (personnaliser les services offerts par le concepteur de l'assistant, mettre en œuvre des campagnes publicitaires plus ciblées, offrir des propositions commerciales plus fines, etc.). Par application du principe de transparence, une telle utilisation des données doit nécessairement être portée à la connaissance des personnes concernées et ne peut en aucun cas se faire à leur insu. Il s'agira d'un nouveau traitement distinct de celui nécessaire au fonctionnement du service, ne pouvant reposer sur la base légale de l'exécution du contrat comme ce dernier et dont la conformité au RGPD devra être analysée de manière séparée.

Il convient également de faire application des dispositions de la loi Informatique et Liberté transposant la directive européenne ePrivacy (Directive 2002/58/CE). En effet, l'article 5(3) de cette directive, transposé à l'article 82 de la loi Informatique et Liberté, prévoit que toute opération de lecture ou écriture sur le terminal d'un utilisateur, au travers d'un réseau de télécommunication ouvert au public, ne peut avoir lieu qu'avec le consentement de celui-ci, à moins que cette opération ne soit strictement nécessaire à la fourniture d'un service explicitement demandé par l'utilisateur ou bien qu'elle vise explicitement à effectuer la transmission d'une communication électronique.

La définition du « terminal de l'utilisateur » englobe des dispositifs tels que des assistants vocaux. Si les données inscrites ou stockées même temporairement (identifiant, enregistrements vocaux) dans le terminal sont accédées depuis des serveurs distants pour alimenter un profilage publicitaire, le recueil de son consentement est nécessaire.

Dès lors, en plus de la question de la conformité du traitement général aux dispositions du RGPD, il convient de se poser celle des finalités de l'opération particulière de lecture des données dans l'assistant. Ainsi, si celles-ci sont nécessaires à la fourniture du service explicitement demandé par l'utilisateur, le consentement n'est pas nécessaire. À l'inverse, si ces opérations visent à enrichir ou créer un profil publicitaire, par nature non nécessaire à la fourniture du service, alors le consentement de l'utilisateur doit être recueilli pour cette finalité particulière.

Il est à noter que les dispositifs installés dans un domicile privé, dont le fonctionnement n'entraîne pas de transmission des données vers l'extérieur (un assistant vocal qui réaliserait localement toutes les opérations, sans nécessiter un échange avec un serveur distant et donc de transmission de données vers un responsable de traitement) peuvent potentiellement bénéficier de l'exemption domestique, comme en dispose l'article 2(2.c) du RGPD.

De tels dispositifs, comme par exemple un assistant vocal non-connecté qui permettrait seulement à un utilisateur de mettre en marche ou d'arrêter un équipement électro-ménager, échapperaient à l'application du RGPD. En privilégiant ce type d'architecture, les fabricants d'assistants vocaux peuvent réduire le risque de violations des données personnelles, tout en allégeant leurs obligations juridiques.

Étape 2 :
Choisir les données collectées et les durées de conservation

Appliquer les principes d'exactitude, de proportionnalité et de minimisation des données

Conformément au principe de minimisation des données, seules les informations strictement nécessaires à la fourniture du service doivent être traitées. L'utilisation des fonctionnalités offertes par l'assistant nécessite évidemment le traitement de la voix pour la détection du mot-clé, la transcription automatique et l'analyse et interprétation de la commande. De plus, si l'assistant vocal est connecté à un compte, seules les données indispensables au fonctionnement du compte et à son interaction avec l'assistant doivent être traitées. Il peut s'agir de données d'identification (nom, prénom) sachant qu'un pseudonyme peut être suffisant, ainsi que des données d'authentification (identifiant et mot de passe). Dans ce cas, les enregistrements et retranscriptions des commandes prononcées par les personnes dans l'environnement de l'assistant sont parfois accessibles depuis l'historique de l'espace utilisateur, qui retrace toutes les recherches et questions enregistrées par l'assistant. Des informations techniques (adresse IP par exemple) ou associées à l'équipement afin de vérifier sa conformité (numéro de série) sont également susceptibles d'être collectées, par exemple à des fins de mises à jour du système ou de sécurité du produit. Certains services optionnels peuvent nécessiter des informations supplémentaires – par exemple, l'utilisateur peut décider de communiquer son code postal pour bénéficier d'informations météorologiques ou de circulation pertinentes.

D'une manière générale, les requêtes adressées par l'utilisateur à son assistant ne doivent pas être utilisées pour déduire des informations la concernant susceptibles de relever de la catégorie des données dites « sensibles » : prétendue origine raciale ou ethnique, opinions politiques, convictions religieuses ou philosophiques ou appartenance syndicale, données concernant la vie ou l'orientation sexuelle. En particulier, toute inférence du type « a recherché les horaires de messe d'une commune = catholique ou croyant », catégorisation ou création de segments sur la base de telles données, aux fins de réaliser un profil doit, avant toute chose, servir une finalité dont la légitimité devra être démontrée et, en tout état de cause, nécessitera préalablement le recueil du consentement explicite, spécifique et informé de l'utilisateur.

S'assurer de la protection des données sensibles : le cas particulier des données biométriques

Certains assistants vocaux proposent à l'utilisateur une option lui permettant de s'identifier à partir de sa voix afin d'accéder à un service différencié des autres utilisateurs, comme, par exemple, les autres membres du foyer. Ainsi, même si plusieurs personnes sont associées à un même appareil, elles auront chacune accès aux informations les concernant (messagerie électronique, agenda, compte client, etc.)

En pratique, une telle fonctionnalité suppose de réaliser la reconnaissance du locuteur, c'est-à-dire d'appliquer un traitement biométrique sur la voix de l'utilisateur. Ainsi, des échantillons de sa voix sont collectés pour créer un modèle ou gabarit biométrique l'identifiant de manière unique afin de le reconnaître lors de toute sollicitation ultérieure de l'assistant.

Utilisées pour identifier une personne, les données biométriques telles que le gabarit de la voix sont qualifiées de données sensibles au sens de la législation en matière de protection des données (voir Chapitre I.1 *La spécificité de la voix*). Le RGPD encadre strictement la reconnaissance vocale biométrique en interdisant par principe de tels traitements, sauf dans certains cas particuliers. Il s'agit notamment du cas où l'utilisateur y a explicitement consenti, de manière totalement libre et informée. Afin de permettre à l'utilisateur de disposer d'une véritable liberté de choix, le concepteur de l'assistant devra notamment offrir un mode d'authentification ou d'identification alternatif à la biométrie, sans contrainte additionnelle.

En application du principe de protection des données dès la conception et par défaut, la CNIL, tout comme le Comité européen à la protection des données personnelles (CEPD)[133], recommande le stockage des données biométriques sur un support placé sous le contrôle exclusif de l'utilisateur (qui peut être en l'espèce l'objet « assistant vocal »), à la main de ce dernier et sécurisé de manière adéquate[134]. Ce mode de fonctionnement devrait être toujours privilégié par rapport à un mode de conser-

133 - CEPD, Guidelines 3/2019 on processing of personal data through video devices, 10/07/2019, https://edpb.europa.eu/our-work-tools/public-consultations/2019/guidelines-32019-processing-personal-data-through-video_en
134 - CNIL, Biométrie dans les smartphones des particuliers : application du cadre de protection des données, https://www.cnil.fr/fr/biometrie-dans-les-smartphones-des-particuliers-application-du-cadre-de-protection-des-donnees

vation des informations biométriques (gabarit de la voix) sur un serveur distant.

Par ailleurs, détecter la voix du bon locuteur suppose de la comparer avec celle d'autres personnes se trouvant aux environs de l'assistant. En d'autres termes, la fonctionnalité de reconnaissance du locuteur implémentée dans les assistants vocaux peut nécessiter d'enregistrer les données biométriques vocales des personnes s'exprimant au sein du foyer, pour permettre de distinguer parmi ces caractéristiques vocales celles de l'utilisateur souhaitant être reconnu. L'identification biométrique peut donc avoir pour conséquence de soumettre des personnes non informées à un traitement biométrique, en enregistrant leur gabarit pour le comparer à celui de l'utilisateur souhaitant être reconnu. Afin d'éviter une telle collecte de données biométriques à l'insu des personnes concernées tout en permettant à un utilisateur d'être reconnu par l'assistant, des solutions reposant sur les seules données relatives à l'utilisateur doivent être privilégiées. Concrètement, cela suppose qu'à chaque utilisation la reconnaissance biométrique n'est activée qu'à l'initiative de l'utilisateur, et non par une analyse permanente des voix entendues par l'assistant. Ainsi, par exemple, si l'utilisateur souhaite mettre en place une authentification biométrique pour l'accès à certaines données protégées comme son compte bancaire, l'assistant vocal pourrait activer la vérification du locuteur, lorsqu'il lance l'application bancaire uniquement et vérifier son identité de cette façon.

Limiter les durées de conservation des données

Conformément au principe de limitation de la conservation (article 5(1.e) du RGPD), les données ne doivent pas être conservées au-delà de la durée nécessaire à la fourniture du service. Par conséquent, les informations devraient être supprimées par le responsable de traitement dès lors que le service a été rendu, par exemple, dès que la réponse à la question posée a été émise par l'assistant.

Parfois, le fournisseur du service offre la possibilité à l'utilisateur de conserver l'historique de ses requêtes sur son compte. Dans cette hypothèse, les données conservées pour être mises à disposition de l'utilisateur ne doivent pas être traitées pour une autre finalité que celle de fourniture du service. Dans le cas contraire, il s'agira d'un nouveau traitement de données, dont la licéité devra de nouveau être préalablement évaluée au regard des règles du RGPD (voir Étape 1). Si l'historique n'est conservé que pour être mis à disposition de l'utilisateur, ce dernier doit avoir la faculté de le supprimer à tout moment.

Lorsque les données ne sont plus nécessaires à la fourniture du service ou lorsque l'utilisateur supprime son compte, le responsable de traitement doit à son tour supprimer définitivement les informations qu'il détient sur l'utilisateur, sous réserve d'éventuelles obligations légales lui imposant de les archiver. La suppression des données permet de réduire les potentiels risques de détournement de finalité. Par exemple, les données collectées pour la fourniture du service pourraient être utilisées pour d'autres finalités telles que la prospection commerciale, l'amélioration du service ou encore l'enrichissement du profil client sans que l'utilisateur n'en soit conscient ou que les dispositions applicables soient respectées (par exemple l'information des utilisateurs, le consentement des personnes s'agissant de certaines finalités, etc.).

Étape 3 :
Informer les personnes et garantir leurs droits

Mettre en œuvre les principes d'information et de transparence

Il existe un seul responsable de traitement sur qui pèsent l'obligation d'information et la gestion des demandes d'exercice des droits des utilisateurs. Dans ce cas d'usage, il revient au concepteur de l'assistant vocal de veiller à délivrer l'information préalable dans des termes clairs et simples avant la mise en œuvre du traitement et au plus tard au moment de la collecte des données (donc au démarrage de l'assistant). Différents supports peuvent être utilisés par les fournisseurs de ces services, de manière cumulative : notice d'usage de l'assistant, pré-enregistrement de messages vocaux explicatifs sur ce dernier qui sont émis jusqu'à confirmation de leur prise en compte par l'utilisateur, information lors de la procédure de création de compte, mise à disposition d'une politique de protection de la vie privée depuis un espace en ligne aisément accessible et identifiable (cette dernière modalité pouvant être utilisée en complément d'une information plus directe et préalable via les supports précités), etc. Le responsable de traitement doit notamment expliquer à l'utilisateur l'objectif de la collecte de ses données, lesquelles sont nécessaires au bon fonctionnement de l'assistant vocal, la manière de garder le contrôle sur ces dernières, notamment par l'exercice de ses droits, etc. S'il en a la possibilité, le responsable de traitement peut s'appuyer sur l'usage d'écran compagnon, par exemple au travers d'une application dédiée, pour la présentation de ces informations ainsi que pour le recueil des actions de l'utilisateur (acceptation des CGU et de la politique de confidentialité, paramétrage du dispositif, etc.).

ZOOM SUR...

L'information vocale

L'interface étant principalement vocale, le fournisseur pourrait utilement prévoir la possibilité d'une information vocale par l'assistant conforme au RGPD. Cette modalité d'information est particulièrement pertinente pour les personnes qui recourent à un assistant vocal et sont dans l'impossibilité d'utiliser un support écrit. Afin d'éviter une longue lecture des conditions d'utilisation et de la politique de vie privée, les principaux aspects de ceux-ci peuvent être intégrés dans une présentation générale succincte, puis complétés par des mécanismes de questions/réponses permettant à l'utilisateur d'accéder aux informations qui l'intéressent le plus.

Assurer le respect effectif des droits des personnes

L'utilisateur dispose de droits lui permettant de garder la maîtrise des informations le concernant. Leur existence et la manière de les exercer doivent être portées à sa connaissance par le responsable de traitement.

L'exercice du droit d'accès permet à l'utilisateur, d'une part, de savoir quelles sont les données détenues à son sujet et d'en obtenir la communication dans un format compréhensible et, d'autre part, d'obtenir des informations relatives au traitement de ces données : les finalités pour lesquelles elles sont traitées, les destinataires de celles-ci, leur durée de conservation, etc. La communication de ses données permet notamment à l'utilisateur d'en contrôler l'exactitude et, au besoin, de les faire rectifier ou d'en demander l'effacement (voir *Les notions clés du RGPD*).

Lorsque cela est possible, le responsable peut permettre à l'utilisateur d'accéder directement à ses données. De nombreux concepteurs d'assistants proposent ainsi à l'utilisateur d'accéder à un historique de ses interactions avec l'assistant vocal. Il est à noter qu'un simple renvoi des utilisateurs à un tel historique ne permet pas, de prime abord, au responsable de traitement de répondre à l'ensemble de ses obligations au titre du droit d'accès, les données accessibles ne représentant généralement qu'une part des informations traitées dans le cadre de la fourniture du service.

Si l'utilisateur constate que les informations le concernant ne sont pas exactes, il peut les rectifier depuis son compte utilisateur ou demander leur modification au responsable de traitement ; il peut également demander leur effacement.

Enfin, l'utilisateur doit se voir offrir la possibilité d'exercer son droit à la portabilité sur ses données, dans le cadre de la fourniture du service par l'assistant vocal et lorsque ces données sont collectées sur la base légale du contrat ou du consentement. Ce droit permet à l'utilisateur de récupérer pour son usage personnel, d'une part, les données qu'il a communiquées dans le cadre de la création de son compte utilisateur (nom, prénom, etc.) et, d'autre part, les données produites par l'utilisation du service (par exemple, l'historique des interactions vocales, etc.). En pratique, les données devront être fournies dans un format adapté et documenté, c'est-à-dire interprétable par un ordinateur et sans restriction d'usage. Cela peut par exemple être réalisé en utilisant un format ouvert (XML, JSON, CSV, WAV, etc.), complété par toute métadonnée utile à leur interprétation. Afin de faciliter l'exercice de ce droit, le responsable de traitement peut offrir aux utilisateurs la possibilité de télécharger directement leurs données depuis leur espace utilisateur.

Étape 4 :
Protéger et sécuriser

Garantir la sécurité des données

En veille permanente, les assistants vocaux peuvent s'activer et enregistrer inopinément une conversation dès lors qu'ils pensent détecter le mot d'activation. Il est donc possible que des conversations privées, intimes ou des données confidentielles voire même sensibles soient captées à l'insu de l'utilisateur, telles que les données relatives à l'état de santé d'un membre de la famille ou encore les données bancaires. Cela est d'autant plus vrai lorsque l'assistant vocal est placé au cœur du foyer et accessible à l'ensemble de ses membres ou des personnes extérieures (amis, personnel de ménage, technicien, etc.)

Ces caractéristiques, inhérentes au fonctionnement même de ce type de service, nécessitent la mise en place de mesures de sécurité renforcées, pour protéger l'espace intime des utilisateurs et des tiers. Ces mesures concernent plusieurs aspects que le responsable de traitement est tenu de couvrir : gestion des habilitations, gestion des incidents, sécurisation des serveurs (usage

de mot de passe fort), encadrement de la maintenance et de la destruction des données, gestion des sous-traitants (voir cas d'usage n°2), sécurisation des échanges (chiffrement des communications), etc. Plus de précisions sur ces précautions élémentaires qui doivent être mises en œuvre de façon systématique et sur la façon de les mettre en œuvre peuvent être trouvées dans le guide de sécurité des données personnelles de la CNIL[135].

Renforcer la maîtrise des données et identifier les risques

Une Analyse d'Impact relative à la Protection des Données (AIPD) a pour objectif de construire et de démontrer la mise en œuvre des principes de protection de la vie privée. Avec le RGPD, mener une AIPD est obligatoire si le traitement est susceptible d'engendrer des risques élevés sur les droits et libertés des personnes concernées. Dès lors, de par la sensibilité des données qui peuvent être traitées lors de l'utilisation d'un assistant vocal, l'AIPD peut être un prérequis.

Cela peut être effectivement le cas si le traitement envisagé figure dans la liste des types d'opérations pour lesquelles la CNIL a estimé obligatoire de réaliser une Analyse d'Impact relative à la Protection des Données[136] ou si le traitement remplit au moins deux des neuf critères issus des lignes directrices du G29[137], le rassemblement des autorités de protection des données européennes (remplacé depuis mai 2018 par le Comité européen à la protection des données personnelles ou CEPD). Ces critères sont : l'évaluation/scoring (y compris le profilage), la prise de décision automatique avec effet légal ou similaire, la surveillance systématique, la collecte de données sensibles ou données à caractère hautement personnel, la collecte de données personnelles à large échelle, les croisement de données, le traitement de données de personnes vulnérables (patients, personnes âgées, enfants, etc.), l'usage innovant (utilisation d'une nouvelle technologie), l'exclusion du bénéfice d'un droit/contrat.

ZOOM SUR...

L'AIPD appliquée au domaine des objets connectés

Pour accompagner les entreprises dans cette démarche, la CNIL a produit des guides AIPD[138] prenant en compte les exigences du RGPD. La méthode est conforme aux critères établis dans les lignes directrices du CEPD et est également compatible avec les normes internationales de gestion des risques. De plus, afin de faciliter la mise en œuvre d'AIPD, la CNIL a développé le logiciel PIA[139], disponible en français et en anglais, qui vise à faciliter et accompagner la conduite d'une telle analyse d'impact.

Plus directement opérationnel pour les personnes développant, intégrant ou déployant des assistants vocaux, la CNIL a également publié une version de la méthode PIA appliquée au domaine des objets connectés[140] aussi appelé PIAF – *Privacy Impact Assessment Framework* en anglais.

135 - CNIL, Le guide de sécurité des données personnelles, édition 2018 https://www.cnil.fr/sites/default/files/atoms/files/cnil_guide_securite_personnelle.pdf
136 - CNIL, Liste des types d'opérations de traitement pour lesquelles une analyse d'impact relative à la protection des données est requise, https://www.cnil.fr/sites/default/files/atoms/files/liste-traitements-aipd-requise.pdf
137 - G29, Lignes directrices concernant l'analyse d'impact relative à la protection des données (AIPD) et la manière de déterminer si le traitement est « susceptible d'engendrer un risque élevé » aux fins du règlement (UE) 2016/679, 2017 https://www.cnil.fr/sites/default/files/atoms/files/wp248_rev.01_fr.pdf
138 - CNIL, Les guides AIPD (analyse d'impact relative à la protection des données), https://www.cnil.fr/fr/guides-aipd
139 - CNIL, Outil PIA : téléchargez et installez le logiciel de la CNIL, https://www.cnil.fr/fr/outil-pia-telechargez-et-installez-le-logiciel-de-la-cnil
140 - CNIL, Analyse d'impact relative à la protection des données – Application aux objets connectés, https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-piaf-connectedobjects-fr.pdf

CAS N°2 :
Utiliser une application bancaire via un assistant vocal

Comme indiqué dans le Chapitre I.4 *Quelle(s) stratégie(s) pour les concepteurs d'assistants vocaux ?*, certains concepteurs d'assistants vocaux permettent à des organismes tiers de développer leurs propres applications, directement interrogeables via l'assistant. Ainsi, ces dernières, accessibles depuis un magasin d'applications, se multiplient pour des usages variés. Il est par exemple possible pour l'utilisateur d'installer une application bancaire dédiée qui lui permet d'avoir accès à certains services via son assistant.

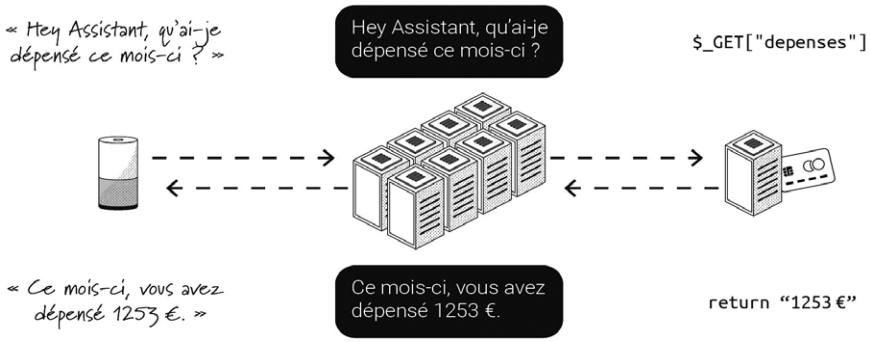


Figure 5
Utilisation d'une application tierce via l'assistant vocal : consultation du compte bancaire

En pratique, comme illustré dans la Figure 5, l'interrogation de sa banque par l'assistant vocal se fait de la manière suivante (plus de détails dans l'infographie page 14) :

- Captation de la requête de l'utilisateur suite à l'activation de l'assistant. La requête précise que l'utilisateur souhaite accéder au service de son établissement bancaire en prononçant la phrase « se connecter à (nom de la banque) » ;
- Transmission aux serveurs du concepteur de l'assistant pour transcription automatique de la parole, interprétation de la commande ;
- Bascule dans l'environnement mis en place par l'établissement bancaire et envoi de la requête vers le système d'information de celle-ci par un système d'interface de programmation d'applications (ou API) ;
- Intégration des informations communiquées par l'établissement bancaire dans la réponse formulée à l'utilisateur par le concepteur de l'assistant.

Pour avoir accès à ce service, l'utilisateur doit disposer dans la majorité des cas d'un compte utilisateur associé à son assistant. Une fois celui-ci créé (voir cas d'usage n°1), il va procéder à l'appairage de son compte bancaire avec l'assistant vocal. Pour cela, il doit, par exemple :

1. S'authentifier sur l'espace client de sa banque ;
2. Autoriser l'appairage entre les deux comptes et valider les conditions générales d'utilisation (CGU) du service ;
3. Une fois l'appairage réalisé et un jeton d'accès (*access token*) délivré, l'utilisateur client de la banque, peut utiliser le service à sa guise, (en prononçant la phrase d'activation de ce service : « se connecter à (nom de la banque) ».

Avertissements sur l'étude de cas n°2

L'étude de ce cas permet de poser les bases juridiques de la relation entre l'utilisateur, le concepteur de l'assistant vocal et le développeur d'application tierce. La question de la répartition des responsabilités y est étudiée en particulier, à travers l'exemple d'une application bancaire. L'analyse développée dans le cas d'usage n°1 sur la création du compte utilisateur pour le paramétrage de l'assistant reste évidemment de rigueur pour cette première étape. Plus généralement, les principes présentés en introduction de ce chapitre et détaillés dans le cas d'usage n°1 demeurent applicables : seules les particularités de ce cas d'usage sont présentées ici.

Étape 1 :
Bien définir le traitement, son responsable et sa base légale

Définir la finalité et le statut des acteurs

Dans ce cas d'usage, deux acteurs interviennent dans le traitement des données à caractère personnel : le concepteur de l'assistant et le développeur de l'application bancaire.

La détermination du rôle des acteurs est un préalable indispensable. Elle permet d'identifier les obligations qui pèsent sur chacune des parties intervenant dans le traitement des données personnelles, de l'information des personnes et la gestion des demandes d'exercice des droits à la notification des failles de sécurité ayant entrainé une violation de données, etc.

Dans le scénario présenté, la banque est le responsable de traitement concernant la fourniture du service puisqu'elle détermine les finalités et les moyens essentiels du traitement liés à l'application permettant d'interagir avec l'assistant. En effet, elle propose une application dédiée qui permet à l'utilisateur, client de l'établissement bancaire, de gérer ses comptes à distance. En outre, elle décide des moyens du traitement même si le sous-traitant, le concepteur de l'assistant, joue un rôle important dans la détermination de ces moyens. À titre d'exemple, il peut opérer la plateforme de développement permettant d'intégrer des applications tierces à l'assistant. Il fixe donc le cadre et les conditions que doivent respecter les développeurs et éditeurs d'applications. En application des dispositions de l'article 28 du RGPD, le sous-traitant doit offrir à son client des garanties suffisantes quant à la mise en œuvre de mesures techniques et organisationnelles appropriées de manière à ce que le traitement réponde aux exigences du présent règlement et garantisse la protection des droits de la personne concernée.

La relation entre le sous-traitant (le concepteur de l'assistant vocal) et l'éditeur de l'application n'est donc pas anodine. En effet, le sous-traitant doit indiquer avec précision dans le contrat les garanties qu'il met en œuvre et ne pas renvoyer à des principes généraux, notamment, en ce qui concerne les mesures de sécurité appliquées, les durées de conservation ou encore les droits des personnes. En pratique, des garanties doivent être apportées quant au traitement des données transitant par l'assistant et nécessaires au bon fonctionnement de l'application bancaire. Le sous-traitant doit en particulier s'engager à ne traiter les données que pour répondre aux besoins de la banque, dans des conditions de sécurité satisfaisantes et à lui notifier dans les plus brefs délais toute éventuelle violation de données. Pour plus d'informations sur les mesures à mettre en place, les parties peuvent se référer au guide du sous-traitant élaboré par la CNIL et disponible sur son site[141].

ZOOM SUR...

Le responsable de traitement et le sous-traitant

Le responsable de traitement est l'entité qui détermine les finalités et les moyens du traitement. Le sous-traitant traite les données pour le compte, sur instruction et sous l'autorité du responsable de traitement. Le sous-traitant peut proposer une solution au responsable de traitement ou même influer sur sa mise en œuvre. Pour autant, il ne décide pas d'y avoir recours et ne traite pas les données pour son propre compte dans le cadre strict de la fourniture du service.

Lorsque l'utilisateur interroge son assistant, sa voix transite sur les serveurs du concepteur de l'assistant vocal pour être retranscrite sous forme de texte et interprétée. Ensuite, la réponse formulée par la banque est enregistrée dans le système d'information du concepteur de l'assistant pour être synthétisée. Dès lors, ce dernier peut accéder aux informations qui circulent à travers ses serveurs afin de répondre à l'interrogation formulée par l'utilisateur. Ces informations ne doivent en aucun cas être exploitées par le concepteur de l'assistant pour son propre compte et à des fins qui lui sont propres, dans la mesure où il agit pour le compte de la banque, en tant que sous-traitant.

141 - CNIL, Règlement européen sur la protection des données : un guide pour accompagner les sous-traitants, https://www.cnil.fr/fr/reglement-europeen-sur-la-protection-des-donnees-un-guide-pour-accompagner-les-sous-traitants

Préciser la base légale du traitement de données

L'application bancaire s'adresse aux utilisateurs de l'assistant vocal déjà clients de l'établissement et ayant conclu une convention de compte avec elle (par exemple, l'ouverture d'un compte bancaire). Il existe donc, indépendamment de l'achat de l'assistant, un contrat entre le client et sa banque.

Comme évoqué précédemment, afin d'utiliser l'assistant vocal pour interagir avec la banque, l'utilisateur devra, au préalable, associer son compte bancaire à l'appareil. L'utilisation du service proposé par la banque correspond donc à une fonctionnalité supplémentaire offerte au client pour consulter ses comptes et avoir accès aux services de la banque. Le traitement de données personnelles nécessaire à la fourniture de cette fonctionnalité peut être fondé sur la base légale de l'exécution du contrat auquel l'utilisateur est partie (article 6(1.b) du RGPD).

En revanche, les traitements réalisés sur ces données pour une finalité autre que la fourniture du service demandé et attendu par l'utilisateur doivent faire l'objet d'une analyse distincte et disposer de leur propre base légale.

Étape 2 :
Choisir les données collectées et les durées de conservation

Appliquer les principes d'exactitude, de proportionnalité et de minimisation des données

Lorsque l'utilisateur choisit de recourir aux services proposés par l'application de la banque, les données collectées correspondent d'une part, aux informations renseignées par l'utilisateur dans son espace client bancaire au moment de l'appairage du compte bancaire et d'autre part, aux enregistrements audio, aux retranscriptions textuelles correspondantes et réalisées par les serveurs du concepteur de l'assistant ainsi qu'aux intentions – les tâches que l'utilisateur demande à accomplir – identifiées dans le texte analysé par la solution du concepteur de l'assistant (voir l'infographie page 14).

Une fois les enregistrements audio retranscrits, l'analyse textuelle se fait, comme pour le cas d'usage n°1, sur les serveurs du concepteur de l'assistant. C'est sur ces instances, dans un espace qui lui est réservé, que la banque a configuré la brique logicielle de son application à partir de fonctions de bases mises à disposition par le concepteur. Les tâches que l'utilisateur peut exécuter à partir de l'application bancaire sont définies : en particulier sont spécifiées les commandes permettant de les mettre en œuvre et les réponses à apporter une fois les intentions correspondantes identifiées. Celles-ci peuvent prendre la forme d'une réponse orale, d'une opération à effectuer ou des deux. À noter que lorsque l'utilisateur choisit de recourir aux services proposés par l'application de la banque, les données nécessaires à l'installation du service correspondent à celles exigées pour la création du compte utilisateur (voir le cas d'usage n°1).

Dans l'hypothèse où le sous-traitant souhaite réutiliser les données pour son propre compte et pour une finalité distincte de la fourniture du service (par exemple, à des fins de personnalisation de la publicité ou pour l'amélioration des services de l'assistant via l'écoute des conversations), il aura alors le statut de responsable de traitement et devra justifier d'une base légale pour ce nouveau traitement qui devra être mis en œuvre conformément au RGPD.

Limiter les durées de conservation des données

Conformément au principe de limitation de la conservation des données (article 5(1.e) du RGPD), celles-ci doivent être détenues pendant une durée n'excédant pas celle nécessaire au regard de la finalité pour laquelle elles ont été initialement collectées ou traitées.

Par conséquent, la banque devra supprimer les données (telles que, par exemple, les transcriptions textuelles, les intentions identifiées ainsi que les journaux ou *logs*, horodatage, etc.) une fois qu'elle a répondu à la demande de l'utilisateur, sauf si elle démontre la nécessité de les conserver pour fournir le service ou pour répondre à une obligation légale ou encore pour des besoins probatoires et conformément aux délais de prescription des actions en justice.

De la même manière, le sous-traitant (le concepteur de l'assistant) devra supprimer les réponses fournies par la banque dès lors qu'elles auront été communiquées à l'utilisateur. Cela doit être effectué indépendamment du fait que l'utilisateur supprime ou non son historique d'activités depuis les paramètres du compte utilisateur (voir cas d'usage n°1).

Étape 3 :
Informer les personnes et garantir leurs droits

Mettre en œuvre les principes d'information et de transparence

La banque, en tant que responsable de traitement doit préalablement à la mise en œuvre du traitement et au plus tard au moment de la collecte des données, informer l'utilisateur sur la finalité, les données collectées, les destinataires, etc. Cette information peut, par exemple, être délivrée lorsque l'utilisateur se connecte à son espace client bancaire au moment où il effectue l'appairage de son compte avec l'assistant vocal. Par ailleurs, dans un objectif de clarté et de pédagogie, il pourrait être utile que la banque explique les différentes étapes du traitement depuis la phase de collecte par le sous-traitant jusqu'à la réponse de la banque via l'assistant, en précisant quelles sont les données accédées par chacun des acteurs, pourquoi et pour combien de temps.

Assurer le respect effectif des droits des personnes

Conformément au RGPD, l'utilisateur peut faire valoir l'ensemble de ses droits auprès du responsable de traitement (accès, rectification, portabilité, etc.) – voir le cas d'usage n°1 et *Les notions clés du RGPD* page 48. S'il appartient au responsable de traitement, c'est-à-dire au développeur de l'application, de fournir l'information aux personnes concernées par les opérations de traitement au moment de la collecte des données, le concepteur de l'assistant doit, dans la mesure du possible, aider le responsable de traitement à s'acquitter de son obligation de donner suite aux demandes d'exercice des droits des personnes concernées.

Étape 4 :
Protéger et sécuriser

Garantir la sécurité des données

On peut identifier deux principaux risques liés à la confidentialité des données. Le premier est relatif à l'accès aux données bancaires par des utilisateurs autres que la personne détentrice du compte, tandis que le second concerne l'accès par le concepteur de l'assistant aux données bancaires.

Tout d'abord, tant que l'appairage de l'assistant au compte client de la banque n'est pas réalisé, aucune information n'est transmise aux serveurs du concepteur de l'assistant. Ensuite, il est à noter que l'appairage qui permet à l'utilisateur de relier son espace client bancaire à son assistant vocal, avec la saisie d'un code de sécurité, se fait en une seule étape qui est pérenne. L'utilisateur pourra choisir de mettre fin à l'utilisation du service en révoquant l'autorisation consentie via la suppression du lien entre son assistant et son espace client depuis les paramètres de l'application bancaire. Ainsi, dès lors que l'appairage a été effectué, toute personne ayant accès à l'assistant peut l'interroger sur les opérations bancaires du bénéficiaire sans qu'il n'y ait besoin de saisir à nouveau le code de sécurité. Par exemple, on peut imaginer qu'une personne de l'entourage de l'utilisateur client de la banque, puisse poser une question à l'assistant sur le détail des trois dernières opérations bancaires alors même qu'elle n'est pas titulaire du compte.

Il est donc souhaitable d'instaurer une durée limitée de l'appairage entre les comptes et l'assistant, des notifications régulières et des rappels des moyens de révocation des jetons d'accès (ou *access token*) de manière à protéger sur la durée les échanges entre les serveurs du concepteur de l'assistant et la banque. En outre, une authentification à deux facteurs, préalable à la mise en relation avec l'application bancaire peut permettre à l'utilisateur de protéger les accès à ses données bancaires. En l'absence d'une telle authentification, et compte tenu de la sensibilité des données bancaires et de la confidentialité de ces informations, il est conseillé aux utilisateurs d'éteindre leur assistant dès lors que des personnes extérieures se trouvent à proximité de celui-ci. Cette mesure de précaution qui vise à éviter l'enregistrement inopiné des conversations permettra également de limiter l'interrogation de l'assistant sur des sujets confidentiels (par exemple, « quel est le solde du compte bancaire ? »).

La relation avec le sous-traitant doit être encadrée de manière stricte. Il est nécessaire de documenter les moyens permettant d'assurer l'effectivité des garanties offertes par le sous-traitant en matière de protection des données : chiffrement des données, chiffrement des transmissions, traçabilité, etc. Le contrat avec le sous-traitant doit notamment définir l'objet, la durée, la finalité du traitement et les obligations des parties relatives à la confidentialité, aux conditions de destruction des données en fin de contrat, à la notification des incidents, etc.

ZOOM SUR...

La confidentialité des données bancaires

Il est possible de distinguer plusieurs types d'opérations bancaires qui nécessitent des niveaux de sécurité différents, allant d'une authentification simple (informations générales sur les comptes) à une authentification beaucoup plus forte (demande de relevé d'identité bancaire, opposition à carte bancaire, virement, etc.), en fonction du niveau de risque présenté par l'opération. Pour garantir la confidentialité des données bancaires lorsque l'utilisateur interroge sa banque, le responsable de traitement devrait distinguer les opérations de gestion courante qui nécessitent une simple vérification de l'identité du client des opérations plus sensibles qui entraîneront des mesures de sécurités adaptées. Par exemple, un virement doit être confirmé par un autre facteur (par exemple, l'envoi d'un secret par SMS ou par mail) permettant l'authentification du client.

CAS N°3 :
Réutiliser les données collectées par l'assistant vocal à des fins d'amélioration du service

L'amélioration du service consiste pour le concepteur à parfaire le fonctionnement de son assistant vocal. Il peut pour cela s'agir d'avoir une meilleure visibilité sur les usages du dispositif en mettant en œuvre des statistiques d'utilisation et de fonctionnement mais également de corriger les capacités de détection du mot-clé, de transcription automatique de la parole et de compréhension automatique du langage.

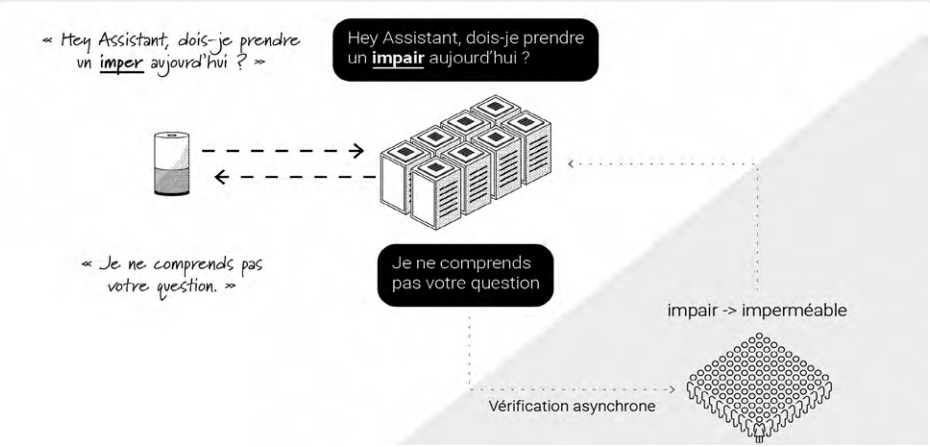


Figure 6
Processus d'amélioration des capacités de compréhension automatique de l'assistant vocal

Ainsi, outre le recueil de données relatives aux usages, sont collectées des données relatives à la façon dont l'utilisateur a interagi avec son assistant vocal. Comme indiqué dans le Chapitre II.2 *Quels enjeux pour les assistants vocaux ?*, les systèmes d'intelligence artificielle sur lesquels reposent les assistants vocaux nécessitent l'utilisation d'exemples d'apprentissage. Or, la catégorisation de ces derniers nécessite une supervision humaine afin d'améliorer les performances des systèmes. Plus précisément, l'amélioration du service consiste à vérifier le fonctionnement opérationnel du système et à remonter les défauts afin que les systèmes soit en mesure de les corriger. Ces opérations sont exécutées par différentes personnes parmi lesquelles on trouve des analystes de la langue, qui s'assurent de la bonne retranscription textuelle en fonction des appareils utilisés, et des analystes de données, qui vérifient le sens des requêtes passées et leur interprétation par l'assistant.

En pratique, comme illustré dans la Figure 6, la réutilisation de données à des fins d'amélioration du service des données collectées par l'assistant vocal se fait de la manière suivante :

- Captation de la requête de l'utilisateur suite à l'activation de l'assistant ;
- Transmission aux serveurs du concepteur de l'assistant pour transcription automatique de la parole, interprétation de la commande ;
- Éventuellement, notification d'un échec à apporter une réponse adaptée à l'utilisateur ;
- Vérification par des personnels de la bonne activation de l'assistant, de la bonne transcription des paroles prononcées, de la bonne exécution de la commande associée à l'ordre, etc. ;
- Annotation de nouveaux exemples d'apprentissage pour amélioration des performances des systèmes d'intelligence artificielle.

Avertissements sur l'étude de cas n°3

L'étude de ce cas permet d'examiner les conditions de licéité de la collecte de données d'interactions avec un assistant vocal à des fins d'amélioration du service. Plus précisément, il permet de questionner la balance à effectuer entre intérêt du concepteur de l'assistant vocal et mesures protectrices de la vie privée de ses utilisateurs. Les modalités d'écoute humaine sont étudiées ici en particulier. L'analyse développée dans le cas d'usage n°1 sur la création du compte utilisateur pour le paramétrage de l'assistant reste évidemment de rigueur. Plus généralement, les principes présentés en introduction de ce chapitre et détaillés dans le cas d'usage n°1 demeurent applicables : seules les particularités de ce cas d'usage sont présentées ici.

Étape 1 :
Bien définir le traitement, son responsable et sa base légale

Définir la finalité et le statut des acteurs

L'objectif poursuivi vise à améliorer les capacités de compréhension vocale de l'assistant pour lui permettre de répondre avec précision aux demandes. Par conséquent, et bien que la finalité d'amélioration du service puisse conduire au traitement de données résultant de l'utilisation d'applications fournies par des tiers, il n'y a qu'un unique responsable de traitement : le concepteur de l'assistant, pour le compte et au bénéfice duquel le traitement est réalisé.

Préciser la base légale du traitement de données

Pour la finalité d'amélioration du service, la base légale du contrat, évoquée dans les cas d'usages n°1 et n°2, n'est pas appropriée. En effet, le traitement des données à caractère personnel n'apparaît pas nécessaire à l'exécution de la prestation attendue par l'utilisateur. En effet, le service demandé par l'utilisateur peut être fourni par l'assistant sans que les données ne soient utilisées à des fins d'amélioration du service. Dans ce cas, seules les bases légales de l'intérêt légitime ou du consentement semblent pouvoir être mobilisées. La démonstration d'un intérêt légitime du responsable du traitement est possible, à moins que les intérêts ou les libertés et droits fondamentaux de la personne concernée ne prévalent. Le traitement doit ainsi relever des attentes raisonnables des personnes et comprendre des garanties lui permettant de garder le contrôle sur ses données. Compte tenu du caractère particulièrement intrusif que constitue l'écoute et l'analyse d'extraits de conversation ou des requêtes des utilisateurs – extraits pouvant au demeurant contenir des données sensibles – la base légale du consentement libre, spécifique et informé devrait être privilégiée, et au-delà des garanties renforcées que doit mettre en œuvre le responsable du traitement, le consentement des personnes concernées constitue la base légale adaptée permettant de préserver le contrôle des personnes sur leurs données (article 6(1.a) du RGPD). L'accès au service rendu par l'assistant ne doit, en tout état de cause, pas être conditionné à l'acceptation par l'utilisateur de l'utilisation de ses données à des fins d'amélioration du service. Par ailleurs, l'accès à des informations collectées par l'assistant vocal, pour d'autres finalités que la fourniture du service demandé par la personne ou la mise en œuvre de communications par voie électronique, nécessite de recueillir le consentement préalable des utilisateurs, conformément à l'article 82 de la loi Informatique et Libertés (voir l'encadré page 52). Pour de plus amples informations sur les conditions de recueil d'un tel consentement, se référer aux lignes directrices de la CNIL sur les cookies et autres traceurs[142].

Étape 2 :
Choisir les données collectées et les durées de conservation

Appliquer les principes d'exactitude, de proportionnalité et de minimisation des données

En application du principe de minimisation des données (article 5(1.c) du RGPD), seules les informations strictement nécessaires pour l'amélioration du service doivent être collectées. Afin de permettre ces améliorations des fonctionnalités de l'assistant vocal, notamment une meilleure compréhension des demandes des utilisateurs, les données collectées correspondent, d'une part, aux enregistrements vocaux des utilisateurs, lorsque l'assistant est allumé et d'autre part, au texte retranscrit.

Limiter les durées de conservation des données

Conformément au principe de limitation de la conservation, les enregistrements vocaux doivent être supprimés dès lors que les correctifs nécessaires pour le bon fonctionnement de l'appareil ont été apportés. Il n'apparaît pas pertinent de conserver les données au-delà de la phase de correction et d'amélioration de l'assistant, d'autant qu'il sera possible au concepteur de l'assistant d'acquérir de nouvelles données pour les améliorations ultérieures. Pour que les informations soient définitivement supprimées, le concepteur de l'assistant doit procéder à leur effacement de son système d'information et, le cas échéant, s'assurer que son sous-traitant en a fait de même (article 28(3.g) du RGPD).

142 - Délibération n° 2019-093 du 4 juillet 2019 portant adoption de lignes directrices relatives à l'application de l'article 82 de la loi du 6 janvier 1978 modifiée aux opérations de lecture ou écriture dans le terminal d'un utilisateur (rectificatif), https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000038783337

Étape 3 :
Informer les personnes et garantir leurs droits

Mettre en œuvre les principes d'information et de transparence

Le responsable de traitement doit informer l'utilisateur au moment de l'achat de l'assistant vocal ou de son installation sur la finalité d'amélioration du service et ce que cette notion recouvre. Si des salariés ou des sous-traitants sont en charge d'écouter les enregistrements vocaux, il devra préciser aux utilisateurs la nature de l'écoute, la durée de conservation des enregistrements, les informations accédées, la base légale, etc. Le responsable de traitement peut réaliser cette information de la manière indiquée dans le cas d'usage n°1 (page 50), par exemple en la faisant figurer dans la notice explicative fournie pour l'installation de l'assistant vocal, ou encore dans les paramètres du compte utilisateur. Ces derniers peuvent permettre d'activer ou non les écoutes réalisées par les salariés/sous-traitants dans le cadre de la finalité d'amélioration du service.

Assurer le respect effectif des droits des personnes

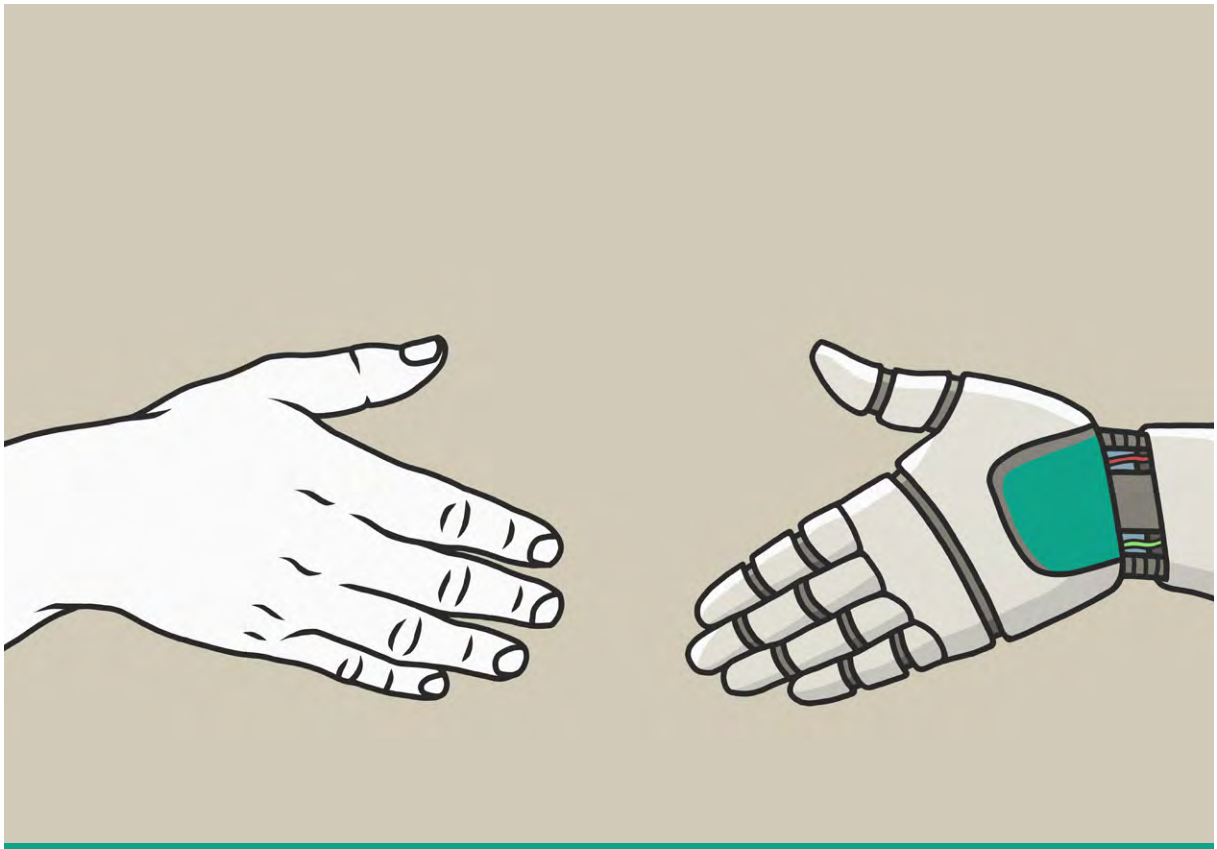
Pour les traitements de données reposant sur le consentement des utilisateurs, le concepteur de l'assistant doit permettre à ce dernier de retirer son consentement à tout moment. Pour cela, une modalité technique simple et équivalente à celle utilisée pour recueillir le consentement doit être mise en œuvre.

La personne concernée doit pouvoir également exercer de manière simple et, si elle le souhaite, par voie électronique, ses droits d'accès, d'effacement (par exemple si elle a retiré son consentement), à la limitation du traitement, et son droit à la portabilité (dans les conditions posées par le RGPD) sur les données utilisées à des fins d'amélioration du service (voir cas n°1 – « Assurer le respect effectif des droits »).

Étape 4 :
Protéger et sécuriser

Garantir la sécurité des données

En premier lieu, comme pour le cas d'usage n°1, les mesures standards de protection des données doivent être mises en œuvre, ainsi que les mesures à prendre lors d'une relation contractuelle avec un sous-traitant relevées dans le cas d'usage n°2. Lorsque le concepteur de l'assistant vocal décide de recourir à des salariés en interne ou à ceux d'une entreprise sous-traitante pour procéder à l'écoute des enregistrements vocaux, il doit mettre en œuvre de nombreuses garanties. Il doit en particulier restreindre l'accès aux données des enregistrements vocaux aux seuls salariés habilités à écouter les conversations dans le cadre de l'analyse des réponses et de la correction de celles-ci à des fins d'amélioration du service. De plus, seules les informations nécessaires doivent être accessibles à ces personnes. Par exemple, il n'est pas nécessaire d'avoir accès aux informations relatives au compte utilisateur dans une optique d'amélioration du moteur de transcription automatique de la parole. La sélection des commandes vocales ainsi que celle des assistants parmi lesquels cette sélection est possible (à travers le consentement de l'utilisateur par exemple) doivent se faire de manière aléatoire. Enfin, les appareils, à partir desquels les échantillons des conversations sont écoutées ne doivent pas être systématiquement les mêmes.



ASSISTANTS VOCAUX, LES BONS RÉFLEXES

Les assistants vocaux présentent d'importants enjeux pour la vie privée, et il est essentiel de rester vigilant. Ainsi, des points d'attention doivent être gardés à l'esprit de toute personne impliquée dans la chaîne de responsabilité d'un assistant vocal, de ses développeurs à ses utilisateurs en passant par ses intégrateurs, éditeurs d'applications ou encore ceux qui feraient le choix de déployer de tels dispositifs dans des lieux ouverts au public ou de passage (tels que des halls d'attente, voitures de location, salles de réunion, etc.).

Un questionnement méthodique et répété est indispensable dans une démarche de *privacy by design*, c'est-à-dire visant à mettre en œuvre les mesures techniques et organisationnelles appropriées pour garantir la protection de la vie privée et des libertés fondamentales dès la conception du projet.

Afin d'instaurer la confiance nécessaire à l'adhésion des utilisateurs aux dispositifs équipés d'assistants vocaux, la CNIL a dégagé quatre principes cardinaux :

1 - Entretenir les frictions désirables : plutôt que de se concentrer sur la mise en œuvre d'une expérience utilisateur absolument sans couture, profiter des moments de frictions (c'est-à-dire des moments de choix, de paramétrages nécessitant l'attention de l'utilisateur) pour présenter la réalité des traitements de données aux utilisateurs de manière adaptée (voir encadré page 69).

2 - Privilégier le local au distant : autant que faire se peut, mettre en œuvre des modalités et capacités de traitement des données directement dans les dispositifs, ce qui confère à l'utilisateur une bonne maîtrise de celles-ci et constitue un facteur de confiance et d'acceptabilité.

3 - Assurer les moyens de contrôle : permettre à l'utilisateur de comprendre et maîtriser les usages qui sont faits de ses données et de paramétrer le fonctionnement du dispositif selon ses choix.

4 - S'adapter au média vocal : se reposer sur des interfaces exclusivement audio soulève d'importants défis en matière de présentation de l'information à l'utilisateur, de recueil de son consentement ou de mise en œuvre de moyens de contrôle. Il est donc nécessaire de mener une réflexion sur les moyens à déployer.

Comme présenté dans la Chapitre III *Cas d'usages : le RGPD en pratique* (page 46), l'utilisation d'un assistant vocal doit répondre aux impératifs de protection des données. Plus spécifiquement, il est nécessaire de s'assurer que tous les grands principes mis en avant dans le RGPD sont bien satisfaits (voir *Les notions clés du RGPD*, page 48) :

Voici donc, à destination des différents publics impliqués dans la chaîne de valeur, quelques bonnes pratiques relatives au développement, au déploiement ou à l'utilisation des assistants vocaux. Celles-ci doivent être appréhendées comme des pistes d'évolution et d'amélioration pour une meilleure protection des utilisateurs et de leurs données personnelles. Par ailleurs, comme indiqué dans le Chapitre I.2 *Assistant vocal, qui es-tu ?*, en fonction des modèles d'affaires et des choix technologiques, certains acteurs peuvent endosser plusieurs combinaisons de rôles et être ainsi concernés par plusieurs des points d'attention mis en avant ici. Le respect de ces conseils ne préjuge pas des décisions que la CNIL pourrait prendre à l'égard des acteurs de cette chaîne, notamment dans le cadre de contrôles ou de procédures contentieuses.

POUR LES CONCEPTEURS D'ASSISTANTS VOCAUX

En se focalisant sur les aspects logiciels, les développeurs d'assistants vocaux sont responsables des implémentations techniques qui vont régir le fonctionnement de ces derniers. Modalités d'activation, choix d'architecture, accès aux données, gestions des enregistrements, spécifications matérielles, etc., c'est par ces choix de conception que se matérialisent les possibilités de l'assistant. Afin de garantir aux utilisateurs maîtrise et contrôle sur leurs données, sept points de vigilance doivent être gardés à l'esprit.

Établir la transparence, fondement de la confiance

Comme pour tout traitement de données personnelles, le RGPD impose un devoir d'information des personnes concernées par les traitements mis en œuvre par des assistants vocaux (voir *Les notions clés du RGPD*, page 48).

Plus encore que pour les services accessibles en ligne, l'usage d'un assistant vocal impose aux utilisateurs de faire confiance à un dispositif qui donne parfois peu d'information sur son fonctionnement et dont les modalités de maîtrise ne sont aujourd'hui pas aussi intégrées que celles qui s'exercent sur un ordinateur ou un smartphone.

Nos conseils

- Être transparent sur le fonctionnement de l'assistant vocal et, notamment, sur les différentes étapes du traitement depuis la phase de collecte en passant par la retranscription de la voix sous forme de texte pour analyse et réponse à l'utilisateur.
- Délivrer cette information avant l'achat du dispositif, soit en portant les mentions d'information sur l'emballage, soit en mettant à disposition des futurs clients une note d'information.
- Rappeler cette information à la première utilisation du dispositif, qui pourrait être également proposée en version audio.
- Si de nombreuses informations doivent être données, prévoir une information par strates permettant de prioriser les éléments à présenter.
- Penser les interfaces pour que les utilisateurs puissent naviguer aisément dans les différentes strates d'information et trouver celles dont ils ont besoin à tout moment (lors de l'installation ou ultérieurement).
- Prévoir une présentation orale et compréhensible des conditions d'utilisation et des règles de protection de la vie privée, accessible en interrogeant l'assistant.
- Permettre à l'utilisateur de poser des questions sur les traitements de données personnelles de l'assistant vocal et fournir des réponses claires par oral.
- Si le concepteur a recours à une utilisation ultérieure des données destinée à améliorer ses propres services – par exemple en employant des personnes pour réaliser des écoutes et annotations des conversations enregistrées – informer spécifiquement et clairement l'utilisateur de cet usage, et indiquer dans l'interface de gestion de l'assistant les commandes et enregistrements qui ont fait l'objet d'une telle utilisation.

- Mettre par défaut le paramétrage du dispositif dans son fonctionnement le plus protecteur de la vie privée pour son utilisateur.
- Si l'assistant vocal nécessite des évolutions logicielles et/ou mises à jour importantes, prévoir un contact avec l'utilisateur pour l'en avertir et lui préciser la nature des changements et leurs conséquences.
- Si le concepteur de l'assistant fournit également un kit de développement d'application (ou SDK pour *Software Development Kit*), intégrer des fonctionnalités et outils logiciels permettant aux développeurs d'applications tierces d'appliquer l'impératif de transparence.

ZOOM SUR...

Données et design, pour des interfaces respectueuses de la vie privée

Avec la publication de son Cahier IP n°6 *La forme des choix*, la CNIL s'est intéressée à promouvoir l'émergence d'un design des interfaces plus responsable et respectueux des principes de protection des données[143]. À l'instar des questions juridiques et techniques, le design des interfaces doit désormais être au centre des préoccupations du régulateur, tout comme il est déjà au cœur des relations entre les individus et les fournisseurs de services.

À la suite de cette publication, la plateforme Données & Design a été lancée[144]. Celle-ci vise à créer des opportunités de collaboration et des espaces d'échange entre des designers pour co-construire des parcours respectueux de la vie privée. L'objectif est d'intégrer concrètement ces réflexions dans le travail quotidien des designers afin de les aider à argumenter leurs choix et à travailler en plus proche collaboration avec d'autres fonctions (chefs de produits, chefs de projets, départements juridiques, etc.) sur la protection des données personnelles.

Divers contenus expliquant et illustrant les points de la réglementation sur lesquels les designers peuvent agir sont mis à disposition. Dans les faits, la plateforme Données & Design est structurée autour d'approches complémentaires relatives à l'explication de concepts clés du RGPD (information des personnes, consentement et exercice des droits), à la mise à disposition d'études de cas et la création d'espaces d'échanges sur ces questions tant en ligne que lors de rencontres physiques. Si les travaux Données & Design ne s'adressent pas spécifiquement aux interfaces vocales, les éléments y figurant peuvent alimenter la réflexion sur les bonnes pratiques à mettre en œuvre avec des assistants vocaux.

Donner des moyens de contrôle aux utilisateurs

Autre exigence majeure portée par le RGPD, la mise en œuvre de moyens permettant aux personnes de maîtriser les usages qui sont faits de leurs données et d'exercer leurs droits de façon simple et effective. Ces modalités de contrôle et d'exercice doivent être adaptées à l'interface vocale de l'assistant.

143 - LINC, Cahier IP6 *La forme des choix,* janvier 2019, https://linc.cnil.fr/fr/cahier-ip6-la-forme-des-choix-0
144 - https://design.cnil.fr/

Nos conseils

Sur l'architecture logicielle

- Promouvoir des architectures logicielles respectueuses de la vie privée des utilisateurs par construction. Par exemple :
  - Pour les services ne nécessitant pas d'accès distant (réveil, pilotage de lumières, etc.) mettre en œuvre des traitements fonctionnant exclusivement de façon locale.
  - Pour minimiser l'exposition des données personnelles, mettre en œuvre au maximum les principes de l'informatique en périphérie (*edge computing*) pour que ne soit transférées sur des serveurs centralisés que les données strictement nécessaires.

Sur les modalités de paramétrage

- Permettre aux utilisateurs de gérer facilement leurs données (écouter, supprimer, détecter des usages anormaux et le cas échéant, récupérer),
  - à travers un tableau de bord accessible via un écran compagnon ;
  - directement par interrogation de l'assistant par la voix.
- Permettre à l'utilisateur de paramétrer finement les fonctionnalités accessibles via son assistant vocal. Par exemple, mettre à sa disposition une fonctionnalité de suppression automatique des informations dès lors que l'utilisateur a obtenu la réponse à sa demande, ou après un délai qu'il peut fixer.
- Prévoir des interfaces fluides et facile d'accès pour la gestion des applications tierces que celles-ci soient vocales et/ou via un écran compagnon et laisser la possibilité à l'utilisateur de désactiver les services et applications préinstallés.
- Envisager, en fonction de la criticité des applications possibles, la mise en place de modalités de filtrage à destination des jeunes enfants activables par leurs parents.

Sur les modalités d'enregistrement

- Offrir à l'utilisateur un moyen de désactiver physiquement le microphone du dispositif.
- Proposer à l'utilisateur une fonction d'activation manuelle, qui peut soit déclencher l'écoute des instructions soit activer une période définie d'attente du mot-clé d'activation.
- Indiquer à l'utilisateur par un signal sonore le début et la fin des périodes d'enregistrement.
- Proposer à l'utilisateur une commande vocale spécifique de désactivation de l'appareil (par exemple lorsqu'il y a des invités, etc.).
- Penser dès la conception la possibilité d'une utilisation par des personnes en situation de dépendance ou de handicap. Par d'exemple, un signal lumineux indiquant que l'appareil est en mode d'écoute actif n'est pas approprié pour des personnes mal-voyantes.

Sur la gestion des comptes

- Permettre d'associer un ou plusieurs comptes personnels à l'assistant en fonction des utilisations possibles de celui-ci.
- Permettre de ne pas associer de compte ou d'associer un compte générique quand l'assistant est destiné à un lieu collectif ou public ou bien à un usage professionnel.
- Proposer un mode de navigation privée pour les actions ne nécessitant pas de s'authentifier, permettant à un utilisateur d'interagir sans qu'un compte soit associé, ni que soient conservées de traces de ces interactions.
- Si plusieurs comptes personnels sont associés à un même dispositif, mettre en œuvre des moyens d'authentification fiables pour le passage de l'un à l'autre et ainsi prévenir de possibles usurpations d'identité.

S'assurer du bon dimensionnement de la collecte de données

Par les interactions qu'ils ont avec eux, les utilisateurs d'assistants vocaux sont susceptibles de transmettre de nombreuses informations malgré eux. Qui plus est, suivant les modalités d'activation des dispositifs (par exemple suite à la prononciation d'un mot-clé), des enregistrements inopinés peuvent également advenir.

Nos conseils

- Déterminer des durées de conservation distinctes selon le type de données collectées. Par exemple, les données associées au compte utilisateur peuvent ainsi être conservées plus longtemps que les requêtes ponctuelles effectuées auprès de l'assistant vocal.
- Ne pas demander de création de compte utilisateur si l'assistant ne le nécessite pas, par exemple si sa fonction est de fournir des renseignements génériques ou de programmer des actions simples.
- Ne pas conserver les enregistrements occasionnés par une fausse activation ou, a minima, les identifier spécifiquement pour que l'utilisateur en soit averti.

Traiter des données biométriques

Certains assistants vocaux ayant vocation à être déployés dans des environnements partagés, des constructeurs proposent d'y associer des comptes pour chaque utilisateur (par exemple les différents membres d'un foyer). Une possibilité pour passer d'un compte à un autre est de se baser sur l'authentification ou l'identification du locuteur. Toutefois, celles-ci reposent sur l'exploitation de données biométriques - les gabarits ou modèles de voix – qui sont considérées comme des données sensibles au sens du RGPD. Pour rappel, celui-ci interdit le traitement de telles données, sauf certaines exceptions limitativement énumérées (article 9(2) - voir *Les notions clés du RGPD*, page 48).

Il est ainsi indispensable de s'assurer que le traitement des données biométriques est désactivé par défaut et conditionné à l'obtention du consentement explicite de chaque personne dont la voix est susceptible d'être ainsi analysée. Qui plus est, l'utilisateur doit disposer d'un mode d'authentification ou d'identification alternatif ne présentant pas de contraintes additionnelles afin qu'il dispose d'une véritable liberté de choix (voir Chapitre III *Cas d'usages : le RGPD en pratique*, page 46).

Nos conseils

- Pour les assistants vocaux non personnels, c'est-à-dire ceux qui sont utilisables par plus d'une personne ou disposés dans un espace partagé, prévoir un mot-clé spécifique ou une question aux personnes présentes et recueillir ainsi leur consentement pour déclencher une reconnaissance biométrique. Par exemple, l'utilisateur peut dire « authentification » ou bien l'assistant peut demander « souhaitez-vous être identifié ? » et attendre une réponse positive pour activer le traitement biométrique.
- Conserver le gabarit biométrique de l'utilisateur sous son contrôle exclusif et privilégier le stockage sur un support individuel, qui peut être le dispositif embarquant l'assistant.
- Réaliser les opérations d'authentification/identification en local, c'est-à-dire directement dans le dispositif embarquant l'assistant.

Satisfaire l'impératif de sécurité

Le RGPD précise que la protection des données personnelles nécessite de prendre des mesures techniques et organisationnelles appropriées afin de garantir un niveau de sécurité adapté aux risques. L'analyse de ces risques constitue donc une étape capitale qui doit être menée avant la conception de l'assistant vocal. Elle peut notamment prendre la forme d'un processus formalisé connu sous le nom sous le nom d'Analyse d'Impact relative à la Protection des Données (AIPD). Cette démarche, qui est obligatoire dans certains cas, et fortement conseillée dans d'autres, a été précisée par la CNIL dans de nombreux outils et méthodes[145] (voir également l'encadré page 56).

La mise en œuvre de tout traitement de données à caractère personnel implique donc une obligation de sécurité. Outre les mesures génériques qui peuvent être retrouvées dans le Guide sur la sécurité des données personnelles[146] et dans le Guide développeur[147] (présenté dans l'encadré page 75), des bonnes pratiques spécifiques aux assistants vocaux peuvent être précisées.

145 - CNIL, L'analyse d'impact relative à la protection des données (AIPD), https://www.cnil.fr/fr/RGPD-analyse-impact-protection-des-donnees-aipd
146 - CNIL, Le guide de sécurité des données personnelles, édition 2018 https://www.cnil.fr/sites/default/files/atoms/files/cnil_guide_securite_personnelle.pdf
147 - CNIL, La CNIL publie un guide RGPD pour les développeurs, janvier 2020, https://www.cnil.fr/fr/la-cnil-publie-un-guide-rgpd-pour-les-developpeurs

Nos conseils

- Choisir une modalité d'activation de l'assistant proportionnée au niveau de risque des services gérés par l'assistant vocal (par exemple un bouton dans certains cas).
- Mettre en œuvre un paramétrage raisonné de la détection du mot-clé : privilégier un taux bas de fausses acceptations pour éviter les déclenchements inopinés.
- Proposer à l'utilisateur de choisir son mot-clé, avec des conseils sur le choix : le mot-clé doit respecter certains critères (ne pas être utilisé trop fréquemment dans les discussions, ne pas être trop proche d'autres mots, etc.).
- Identifier les applications à risque et proposer pour celles-ci des mesures de sécurité comme l'authentification à deux facteurs (par exemple via une validation à effectuer suite à un envoi de courriel ou de SMS).
- Réaliser une Analyse d'Impact relative à la Protection des Données et la mettre régulièrement à jour afin d'assurer que les mesure techniques et organisationnelles prises sont en adéquation avec les risques que le traitement de données fait peser sur les personnes (pour plus de précisions sur l'AIPD, voir l'encadré page 56).
- Prévoir des mécanismes d'information et d'alerte de l'utilisateur en cas de dysfonctionnement de l'assistant ou d'activités inhabituelles, notamment en cas de violations de données[148].

Organiser son écosystème applicatif

Comme présenté précédemment, certains assistants – et notamment les plus répandus auprès du grand public – se positionnent comme une plateforme pour héberger des applications tierces. Ces modes de fonctionnement doivent s'accompagner de mesures spécifiques et d'une attention renforcée sur le partage des données.

Nos conseils

- Lorsqu'un acteur tiers utilise les ressources technologiques mises à disposition pour le développement de son application, définir contractuellement les règles applicables en matière de confidentialité et de respect de la vie privée de manière suffisamment claire et précise.
- Préciser la chaîne de responsabilité impliquant le concepteur de l'assistant et le développeur de l'application.
- Accompagner les développeurs d'application dans la mise en œuvre d'un service sécurisé, comme des API d'authentification génériques et de présentation d'information adaptée au dispositif.
- Limiter le nombre d'applications disponibles par défaut au strict nécessaire et favoriser l'installation des applications à l'initiative de l'utilisateur, par exemple via un magasin (*store*), plutôt qu'une mise à disposition directe depuis l'assistant, sans sélection préalable ni information spécifique.
- Lorsque l'assistant accède directement à une application tierce, ne partager aucune donnée personnelle avec le tiers sans une information claire de la personne.
- Mettre en œuvre une politique de validation des applications déposées dans le magasin (store) et régulièrement vérifier celui-ci, notamment en surveillant la présence d'applications aux noms très proches d'applications légitimes.
- Offrir à l'utilisateur des outils de contrôle granulaire pour toutes les applications installées et les données accédées par celles-ci, notamment les données sensibles ou révélatrices de sa vie privée (données de santé ou biométriques, géolocalisation, historique de recherche, etc.). Ces contrôles devraient aussi pouvoir être temporaires (accorder un accès une seule fois ou pour une durée limitée).
- Si des protocoles d'autorisation sont mis en œuvre pour permettre à une application tierce d'accéder à un service, s'assurer que les jetons (*tokens*) permettant l'authentification des utilisateurs ont une durée de vie limitée et raisonnable et qu'ils soient facilement révocables.

148 - CNIL, Notifier une violation de données personnelles, https://www.cnil.fr/fr/notifier-une-violation-de-donnees-personnelles

Encadrer l'utilisation des données pour l'amélioration des technologies

Les assistants vocaux, comme d'autres objets connectés, peuvent faire remonter des données personnelles vers leurs concepteurs à des fins d'amélioration des services. Il peut s'agir de données techniques utilisées à des fins statistiques (par exemple, des informations relatives à l'utilisation d'une ampoule connectée et à sa durée de vie). Dans le cas spécifique des assistants vocaux, la question de l'amélioration du produit peut également impliquer des traitements de données issues des commandes vocales, tels que les enregistrements audio ou leurs retranscriptions textuelles.

En effet, ceux-ci mettent en œuvre des algorithmes d'intelligence artificielle dont les performances sont directement corrélées à des jeux de données utilisés pour l'apprentissage de modèles statistiques. Par conséquent, il peut être légitime de souhaiter accéder à des données relatives à l'utilisation en conditions réelles du dispositif pour travailler à son amélioration. En pratique, il peut par exemple s'agir pour le concepteur de l'assistant vocal de recourir à des salariés en interne ou à ceux d'une entreprise sous-traitante pour procéder à l'écoute et annotation des enregistrements vocaux afin que ceux-ci puisse permettre l'amélioration des modèles. Une grande attention doit être portée aux modalités de mise en œuvre de telles utilisations des données.

Il convient ainsi de respecter les droits de personnes concernées en ne traitant aucune donnée pour cette finalité sans s'assurer de leur bonne information et de la base légale de leur traitement Qui plus est, ces personnes doivent être dans la capacité de savoir clairement et à tout moment si leurs données sont utilisées à cette fin et de s'y opposer facilement.

Nos conseils

- Mettre en œuvre des mesures de sécurité fortes : restriction de l'accès aux données des enregistrements vocaux aux seuls salariés habilités à écouter les conversations, authentification forte, mesure de traçabilité renforcé, blocage de l'extraction des enregistrements audio, etc.
- Ne pas fournir d'informations relatives à l'utilisateur de l'assistant en accompagnement de l'enregistrement et en particulier, ne pas mettre en correspondance les enregistrements et fichiers transcrits avec d'autres données susceptibles d'être collectées (identifiants de l'appareil, localisation, comptes associés, etc.).
- Mettre en œuvre des mesures d'altération des informations présentes dans les enregistrements audio en procédant par exemple :
  - À une modification des caractéristiques du locuteur (timbre, pitch, prosodie, etc.) de manière irréversible ;
  - À une suppression/offuscation des informations contenues dans les enregistrements et retranscriptions (noms et prénoms, adresse, etc.).
- Échantillonner le message concerné en plusieurs parties qui seront analysées par des personnes différentes.
- Limiter la durée des écoutes des conversations des utilisateurs à quelques secondes par échantillon.
- S'assurer que les appareils à partir desquels les échantillons des conversations sont écoutées ne sont pas systématiquement les mêmes.
- En cas de recours aux salariés d'une entreprise sous-traitante, prévoir dans le contrat de sous-traitance toutes les garanties nécessaires en matière de sécurité (nécessité d'inclure une clause de confidentialité dans les contrats de travail du personnel concerné, modalités d'accès aux locaux et aux données, processus d'habilitation des personnes, durées de conservation des données, etc.).

POUR LES DÉVELOPPEURS D'APPLICATIONS

Certains assistants – et notamment les plus répandus auprès du grand public – se basent sur des approches de type plateforme pour héberger des applications tierces. En pratique, il est généralement nécessaire de respecter les contraintes de développement imposées par le concepteur de l'assistant. Toutefois, de bonnes pratiques de développement peuvent également être observées.

Mettre en œuvre les principes de *Privacy by design* et de responsabilité

Les développeurs d'applications se voient souvent proposer des kits de développement (*Software Development Kit* ou SDK). Dès lors, il leur faut penser le cycle de vie des données dans un cadre spécifique : Comment circulent-elles ? Où se situent les responsabilités ? Quelles sont les bonnes pratiques à mettre en œuvre ? Le chapitre « Maîtriser vos bibliothèques et vos SDK » du Guide du développeur de la CNIL permet de couvrir les points techniques à vérifier (voir l'encadré ci-contre).

Nos conseils

- Être transparent et expliquer les différentes étapes du traitement depuis la phase de collecte par le concepteur de l'assistant jusqu'à la réponse de l'application, en précisant quelles sont les données accédées par chacun des acteurs, pourquoi et pour combien de temps.
- Ne collecter que les seules données nécessaires à la réalisation de l'application.
- Vérifier que la collecte de données pour l'application ne déclenche pas d'autres collectes de données par le concepteur ou des tiers associés.
- Vérifier régulièrement les fonctionnements des SDK et API ainsi que les données collectées par ces canaux.
- Contrôler et sécuriser les données personnelles transmises de l'application à l'utilisateur par le biais de son assistant vocal en prenant des précautions particulières pour les données personnelles très révélatrices de la vie privée des utilisateurs (données de consommation énergétiques, soldes bancaires, données de santé etc.).
- Définir clairement les termes de contrat et engagements relatifs aux questions de vie privée concernant l'utilisation de ressources mises à disposition par le concepteur de l'assistant et ne pas conclure de contrats génériques ne prenant pas en compte les spécificités des besoins de l'application.
- Préciser de façon claire la chaîne de responsabilité impliquant le concepteur de l'assistant et le développeur de l'application. Sur le périmètre qui relève de la responsabilité de l'entreprise développant son application, mettre en place les obligations du RGPD, notamment l'information des personnes, l'exercice de leurs droits et la sécurité des données.
- Utiliser au mieux les possibilités laissées par le concepteur du SDK pour délivrer une information claire et proposer des mécanismes d'authentification appropriés lors des premiers paramétrages.

ZOOM SUR...

Le Guide RGPD du développeur

Afin d'accompagner les développeurs dans la mise en conformité de projet web ou applicatif, la CNIL a élaboré un guide de bonnes pratiques des développements en open source[149] [150]. Il propose des conseils et des bonnes pratiques, et offre ainsi des clés de compréhension du RGPD utiles pour tous les acteurs, quelle que soit la taille de leur structure.

Ce guide est découpé en 16 fiches thématiques qui couvrent la plupart des besoins des développeurs pour les accompagner à chaque étape de leur projet, de la préparation du développement à la mesure de l'audience. Ces bonnes pratiques, qui n'ont donc pas vocation à couvrir l'ensemble des exigences des règlementations ni à être à prescriptives, apportent un premier niveau de mesures permettant de prendre en compte les problématiques de protection de la vie privée dans les développements informatiques qui ont vocation à être appliquées à tous les projets traitant des données personnelles.

149 - CNIL, Notifier une violation de données personnelles, https://www.cnil.fr/fr/notifier-une-violation-de-donnees-personnelles
150 - https://github.com/LINCnil/Guide-RGPD-du-developpeur

POUR LES INTÉGRATEURS D'ASSISTANTS VOCAUX

S'ils sont des constructions logicielles, les assistants vocaux ont cependant vocation à être incarnés dans des équipements physiques. Ces objets connectés peuvent être très divers : smartphone, véhicule, robot ménager, enceinte de salon, jouet pour enfant, etc.

Dans certains cas, les développeurs et intégrateurs d'assistants vocaux peuvent être la même entité, mais ce n'est pas nécessairement le cas. Si les développeurs d'assistants vocaux se concentrent sur les aspects logiciels (tout en apportant des indications en termes de spécifications requises), les intégrateurs se focalisent eux sur les contraintes matérielles. En tout état de cause, il convient de noter que les cas d'usage sont nombreux, les contextes d'utilisation multiples et les publics visés différents en fonction des applications. Les conseils proposés ici sont donc génériques, mais peuvent être enrichis en fonction des modalités d'usages. Dans tous les cas, il convient d'être particulièrement attentif aux modalités d'information des personnes.

Établir la transparence, fondement de la confiance

Comme pour tout traitement de données personnelles, le RGPD impose une information des personnes concernées par les traitements mis en œuvre par des assistants vocaux. Plus encore que pour les services accessibles en ligne, l'usage d'un assistant vocal impose aux utilisateurs de faire confiance à un dispositif qui donne peu d'information sur son fonctionnement et dont les modalités de maîtrise ne sont aujourd'hui pas aussi intégrées que celles qui s'exercent sur un ordinateur ou un smartphone.

Nos conseils

- Vérifier que les conditions d'information et de transparence prévues par le concepteur de l'assistant sont bien satisfaisantes pour permettre de traiter les données des personnes conformément à la législation (voir les conseils pour les concepteurs d'assistants ci-dessus, page 68).
- Mettre en œuvre l'information prévue et, le cas échéant, une information complémentaire adéquate.
- S'il est envisagé, dans une évolution logicielle à venir, de doter un objet d'un assistant vocal, bien préciser dans les spécifications de l'équipement si des capacités de production de son (haut-parleur), d'écoute (microphone) et de traitement (processeur) sont embarquées.
- Informer spécifiquement les utilisateurs quand une fonctionnalité d'assistant vocal vient à être déployée sur un équipement qui ne le proposait pas initialement, et leur permettre de continuer à bénéficier d'un équipement pleinement fonctionnel sans activer l'assistant vocal s'ils le souhaitent.

Donner des moyens de contrôle aux utilisateurs

Autre exigence majeure portée par le RGPD, la mise en œuvre de moyens permettant aux personnes de maîtriser les usages qui sont faits de leurs données et d'exercer leurs droits de façon simple et effective. Ces modalités de contrôle et d'exercice doivent être adaptées à l'interface vocale de l'assistant.

Nos conseils

- Réfléchir préalablement à l'intérêt et aux attendus de l'intégration d'un assistant vocal dans l'équipement en question.
- Si un tel choix est effectivement pertinent, choisir l'assistant à intégrer en fonction des objectifs poursuivis et de l'impératif de protection de la vie privée.
- Laisser le choix à l'utilisateur d'utiliser ou non l'assistant intégré à son équipement si celui-ci n'est pas absolument nécessaire au service proposé, tout en continuant à bénéficier d'un équipement pleinement fonctionnel.

- Mettre en œuvre un bouton de désactivation physique du microphone (électriquement non alimenté).

Satisfaire l'impératif de sécurité

De la même manière que pour les concepteurs d'assistants vocaux, le RGPD précise que la protection des données personnelles nécessite de prendre des mesures techniques et organisationnelles appropriées afin de garantir un niveau de sécurité adapté aux risques. La mise en œuvre de tout traitement de données à caractère personnel implique donc une obligation de sécurité. Outre les mesures génériques qui peuvent être retrouvées dans le Guide sur la sécurité des données personnelles[151] et dans le Guide développeur[152] (présenté dans l'encadré page 75), des bonnes pratiques spécifiques aux assistants vocaux peuvent être précisées.

Nos conseils

- Déployer les assistants vocaux sur des équipements mis à jour et correctement sécurisés (voir par exemple l'encadré sur les jouets connectés, ci-après).
- Privilégier les assistants dont le fonctionnement est maîtrisable, c'est-à-dire pour lesquels il est possible d'agir sur l'ensemble du paramétrage technique et de la sélection des fonctionnalités.
- Éviter les assistants susceptibles de transmettre des données à un tiers sans connaître les conditions de traitement des données par celui-ci.
- Éviter les assistants opérés par des acteurs qui réutilisent les données pour leur propre compte ou veiller à encadrer contractuellement les traitements réalisés par le concepteur de l'assistant.
- Réaliser une Analyse d'Impact relative à la Protection des Données et la mettre régulièrement à jour afin d'assurer que les mesures techniques et organisationnelles prises sont en adéquation avec les risques que le traitement de données fait peser sur les personnes (pour plus de précisions sur l'AIPD, voir l'encadré page 56).

ZOOM SUR...

Les jouets connectés pas toujours sécurisés

En 2017, la CNIL a effectué des missions de vérification sur deux jouets connectés. Ces jouets, équipés d'un microphone et d'un haut-parleur, répondent aux questions des enfants sur des sujets divers tels que les fées et les dinosaures. La réponse est extraite d'Internet et donnée à l'enfant par l'intermédiaire de ces objets.

Les contrôles réalisés ont permis de relever que la société qui commercialise ces jouets collecte par leur intermédiaire une multitude d'informations personnelles sur les enfants et leur entourage, notamment leur voix et le contenu des conversations échangées. Plus encore, il a été constaté que le défaut de sécurisation des jouets permet à toute personne possédant un dispositif équipé d'un système de communication Bluetooth de s'y connecter, à l'insu des enfants et des adultes les entourant, et d'avoir ainsi accès aux discussions échangées dans un cercle familial ou amical.

Au vu de ces éléments, la Présidente de la CNIL a considéré que les traitements mis en œuvre n'étaient pas conformes à la loi Informatique et Libertés, en raison du non-respect de la vie privée des personnes et de l'absence d'information des personnes concernées, et a décidé en conséquence de mettre en demeure le responsable de traitement d'adopter des mesures correctrices sous un délai de deux mois. Cette mise en demeure a été rendue publique en décembre 2017.

151 - CNIL, Le guide de sécurité des données personnelles, édition 2018 https://www.cnil.fr/sites/default/files/atoms/files/cnil_guide_securite_personnelle.pdf
152 - CNIL, La CNIL publie un guide RGPD pour les développeurs, janvier 2020, https://www.cnil.fr/fr/la-cnil-publie-un-guide-rgpd-pour-les-developpeurs

POUR LES ORGANISMES SOUHAITANT DÉPLOYER DES ASSISTANTS VOCAUX

Comme les stratégies des différents concepteurs d'assistants vocaux semblent l'indiquer, une évolution tendancielle forte est le déploiement de ces assistants dans des environnements de plus en plus partagés. Les initiatives en ce sens sont nombreuses : partenariats avec des chaînes hôtelières, intégration de série dans des véhicules dont certains seront loués, ligne de produits à destination du monde de l'entreprise, etc.

En tout état de cause, il convient de noter que les cas d'usage sont nombreux, les contextes d'utilisations multiples et les publics visés différents en fonction des lieux d'implémentation de ces technologies. Comme développé dans le Chapitre II.2 *Quels enjeux pour les assistants vocaux ?*, les déploiements dans des lieux ouverts ou de passage soulèvent de très nombreuses questions auxquelles il convient de répondre avant toute mise en œuvre. Cela est en particulier le cas pour les lieux publics, non traités ici, pour lesquels un encadrement juridique spécifique est nécessaire. Les conseils proposés ici sont donc génériques, mais peuvent être enrichis en fonction des usages. Une utilisation professionnelle n'emporte ainsi pas les mêmes risques et obligations qu'une plus ludique, qui diffère encore de celle qui serait faite par des personnes en situation de dépendance. Dans tous les cas, il convient d'être particulièrement attentif aux modalités d'information des personnes.

Établir la transparence, fondement de la confiance

Comme pour tout traitement de données personnelles, le RGPD impose un devoir d'information des personnes concernées par les traitements mis en œuvre par des assistants vocaux. Plus encore que pour les services accessibles en ligne, l'usage d'un assistant vocal impose aux utilisateurs de faire confiance à un dispositif qui donne peu d'information sur son fonctionnement et dont les modalités de maîtrise ne sont aujourd'hui pas aussi intégrées que celles qui s'exercent sur un ordinateur ou un smartphone. Dans le cas d'un environnement professionnel, l'information doit être particulièrement adaptée au contexte de mise en place de l'assistant. Ainsi des contraintes particulières peuvent s'appliquer, comme la consultation des instances représentatives du personnel ou, plus largement, le droit du travail, ou la prise en compte de publics spécifiques.

Nos conseils

- Informer toutes les personnes susceptibles de voir leurs conversations enregistrées par le dispositif.
- Le cas échéant, prévoir une modalité de recueil du consentement des personnes et un mode de fonctionnement alternatif, nécessaire pour un consentement libre.
- Lorsqu'il est embarqué dans un dispositif dédié, positionner l'assistant vocal à un endroit où il sera bien en évidence et visible de tous.
- Déterminer si certaines catégories de personnes vulnérables (personnes âgées, dépendantes, enfants, etc.) sont susceptibles d'être concernées et prendre les mesures nécessaires (voir l'encadré ci-contre).

ZOOM SUR...

La mise en œuvre d'un assistant vocal à destination de personnes dépendantes

Dans le cas d'assistants à destination de personnes en situation de perte d'autonomie, la chaîne de responsabilité des assistants vocaux implique, outre les utilisateurs, concepteurs, développeurs d'applications tierces, un tiers aidant. En effet, qu'il s'agisse d'un membre de famille, un aide à domicile, un assistant social, un prestataire de services indépendant, ou encore d'un représentant du fabricant du matériel, son intervention peut s'avérer nécessaire pour mettre en place et configurer l'appareil. Selon les cas, certaines de ces parties prenantes peuvent être liées à l'utilisateur final par un contrat de vente de matériel ou d'abonnement au service, de services d'installation et de maintien, etc. D'autres, notamment des membres de famille, peuvent intervenir matériellement dans le cadre de la mise en place ou du fonctionnement de l'assistant vocal, sans que leur rôle soit formellement défini. Dans ces conditions, la qualification du statut de ces personnes au regard de la réglementation – en tant que personne concernée, responsable de traitement, sous-traitant, etc. – ne pourra être faite qu'au cas par cas. S'agissant de certains cas extrêmes de perte d'autonomie, la capacité même des personnes concernées à accomplir des actes juridiques – dont l'acte de consentir au traitement de leurs données – pourrait être problématique. Eu égard à ces considérations, le consentement n'apparait donc pas, en général, être une base légale adaptée.

Les données susceptibles d'être collectées sont les mêmes que celles utilisées classiquement : données d'identité, données d'authentification, données d'informations de contact, etc. Dans le cas de l'intervention d'un tiers aidant (un proche, un aidant à domicile ou un prestataire de service) des garanties spécifiques doivent être mises en place. Cela afin de limiter les risques de violation de données, d'atteinte à la vie privée ou encore à l'usurpation d'identité, dans la mesure où certaines données, notamment celles d'authentification, ont vocation à demeurer confidentielles et connues uniquement de la personne concernée.

Par ailleurs, il est à noter que certaines fonctionnalités ciblant des personnes en situation de perte d'autonomie peuvent nécessiter un traitement de données sensibles (par exemple, un pense-bête pour la prise des médicaments). De même, la manière d'utiliser les assistants vocaux peut être révélatrice de certaines fragilités des utilisateurs (ainsi, l'amplification du son lors d'une conversation téléphonique, l'émission d'un appel d'urgence, ou encore des commandes vocales incohérentes). Enfin, les métadonnées liées à l'utilisation du dispositif peuvent fournir, en cas d'accès par un tiers non-autorisé, des indications sur l'activité physique (par exemple, le fait d'être ou non présent au domicile) de l'utilisateur. L'accès non-autorisé à de telles données est susceptible de générer des risques d'autant plus importants que les utilisateurs de ces dispositifs peuvent souvent vivre en état de relatif isolement et/ou être en situation de perte d'autonomie (et, donc, de fragilisation physique ou cognitive). Aussi une attention importante doit être portée, lors de la conception et fabrication de ces dispositifs, à des mesures visant à en assurer la sécurité. À ce titre, la réalisation d'une Analyse d'Impact relative à la Protection des Données personnelles peut alors être nécessaire (voir encadré page 56).

Donner des moyens de contrôle aux utilisateurs

Autre exigence majeure portée par le RGPD, la mise en œuvre de moyens permettant aux personnes de maîtriser les usages qui sont faits de leurs données et d'exercer leurs droits de façon simple et effective. Ces modalités de contrôle et d'exercice doivent être adaptées à l'interface vocale de l'assistant.

Nos conseils

- Privilégier l'utilisation de dispositifs équipés d'un bouton de désactivation physique du microphone.
- Envisager un moyen d'activation moins incertain que la détection de mot-clé (par exemple par activation d'un bouton physique).
- Laisser la possibilité de couper le micro à la main des personnes.
- Choisir l'assistant à déployer en fonction de ses caractéristiques et spécifications. Par exemple :
  - Modalités de gestion des données ?
  - Existence d'une réutilisation des données ?
  - Mise en œuvre de traitement de données locaux/distants ?
- S'assurer que les utilisateurs disposent bien de moyens leurs permettant d'exercer leurs droits sur leurs données (information, consultation, accès, effacement, opposition), par exemple en conditionnant l'activation du dispositif à la fourniture d'un moyen de contact (adresse courriel par exemple).
- Privilégier le choix d'assistants vocaux proposant un mode de navigation privée pour les actions ne nécessitant pas de s'authentifier et permettant ainsi à un utilisateur d'interagir sans qu'un compte soit associé, ni que soient conservées de traces de ces interactions.
- Configurer l'assistant pour qu'il se réinitialise à brève échéance et qu'aucune donnée ne soit conservée au-delà de l'interaction envisagée, en particulier dans les lieux de passage.

Satisfaire l'impératif de sécurité

De la même manière que pour les concepteurs d'assistants vocaux, le RGPD précise que la protection des données personnelles nécessite de prendre des mesures techniques et organisationnelles appropriées afin de garantir un niveau de sécurité adapté aux risques. La mise en œuvre de tout traitement de données à caractère personnel implique donc une obligation de sécurité. Outre les mesures génériques qui peuvent être retrouvées dans le Guide sur la sécurité des données personnelles[153] et dans et dans le Guide développeur[154] (présenté dans l'encadré page 75), des bonnes pratiques spécifiques aux assistants vocaux peuvent être précisées. En particulier, si l'assistant est rendu accessible dans un lieu ouvert au public ou sur un réseau accessible à un grand nombre d'utilisateurs, la sécurisation de son utilisation appelle des mesures additionnelles.

Nos conseils

- Analyse d'Impact relative à la Protection des Données (AIPD) et la mettre régulièrement à jour afin d'assurer que les mesure techniques et organisationnelles prises sont en adéquation avec les risques que le traitement de données fait peser sur les personnes (voir l'encadré page 56).
- Déployer les assistants vocaux sur des équipements mis à jour et correctement sécurisés (voir l'encadré sur les jouets connectés page 77).
- Choisir avec attention les services qui peuvent être pilotés par l'assistant vocal, identifier ceux potentiellement à risque et contrôler strictement l'administration de l'assistant.
- Être vigilant à n'installer et n'accéder qu'à des applications légitimes, des pirates pouvant créer des applications malveillantes afin de collecter des données ou d'entrer dans le système d'information de l'organisme.
- Dans le cas d'un déploiement en environnement professionnel, prendre en compte dès le stade d'idéation les différents risques qui peuvent peser sur l'organisation : risques sur la vie privée, sur la sécurité des systèmes d'information dont les risques sur la confidentialité de certaines informations sensibles ou stratégiques pour l'organisme, etc. et prendre des mesures en conséquence.

153 - CNIL, *Le guide de sécurité des données personnelles*, édition 2018 https://www.cnil.fr/sites/default/files/atoms/files/cnil_guide_securite_personnelle.pdf
154 - CNIL, *La CNIL publie un guide RGPD pour les développeurs*, janvier 2020, https://www.cnil.fr/fr/la-cnil-publie-un-guide-rgpd-pour-les-developpeurs

Respecter les droits des salariés

Faire le choix de déployer un assistant vocal dans un environnement professionnel peut être motivé par le souhait de faciliter le quotidien des employés, l'amélioration des outils de travail à leur disposition, etc. Si de tels outils peuvent apparaître légitimes, ils ne doivent toutefois pas conduire à placer les employés sous surveillance constante et permanente. Il convient donc d'encadrer de près la mise en place de tels dispositifs techniques.

Au regard de la jurisprudence de la Cour de cassation en matière sociale, les salariés doivent être informés des périodes pendant lesquelles ils sont susceptibles d'être écoutés ou enregistrés : quand bien même ce n'est pas la finalité de l'installation d'un assistant vocal dans un espace de travail, le risque de détournement des enregistrements réside.

Nos conseils

- Informer les salariés individuellement (courrier électronique, annexe au contrat de travail si nécessaire, etc.) et collectivement (via l'information et la consultation, le cas échéant, des instances représentatives du personnel) préalablement au déploiement d'un ou plusieurs assistants vocaux.
- L'information doit notamment préciser où peuvent être déployés ces dispositifs (salle de réunion, bureau d'un salarié, etc.), qui peut y accéder, à quelles fins, pendant combien de temps, et quels sont les droits dont disposent à cet égard les salariés.
- Définir de façon claire la chaîne de responsabilité impliquant l'employeur, le concepteur de l'assistant et le développeur de l'application.
- Lorsque qu'il est embarqué dans un dispositif dédié, positionner l'assistant vocal à un endroit où il sera bien en évidence et visible de tous.
- Encadrer l'utilisation de tels dispositifs (endroits où ils peuvent être déployés, les conditions de leur mise en marche et arrêt, le processus de consultation, par l'employeur, des données collectées ou générées par les dispositifs, les sanctions éventuelles en cas de non-respect des consignes, etc.). Ces précisions peuvent notamment être inclues dans le règlement intérieur ou la charte informatique de l'entreprise.
- Prévoir des mesures de suppression des données personnelles, comptes d'utilisateur, etc., pour les salariés dont le contrat du travail prendrait fin.

POUR LES UTILISATEURS

Modalités d'activation et d'information, services et usages disponibles, mesures de sécurité… choisir d'utiliser un assistant vocal n'est pas anodin. Il convient d'être conscient des enjeux posés par ces dispositifs. Cinq points de vigilance sont à noter pour les utilisateurs.

Veiller à la confidentialité des échanges

En veille permanente, l'assistant vocal peut s'activer et enregistrer inopinément une conversation dès lors qu'il croit avoir détecté le mot-clé. Une fois enregistrées, les interactions sont susceptibles d'être écoutées par des personnes, salariés ou prestataires de la société fournissant l'assistant vocal, en vue d'améliorer les différents algorithmes mis en œuvre (détection du mot-clé, transcription automatique, compréhension du langage, etc.). Choisir de placer un tel dispositif au cœur de son foyer ou de son véhicule implique donc des responsabilités envers les différentes personnes dont les données personnelles sont susceptibles d'être traitées.

Nos conseils

Sur le choix du dispositif à utiliser

- Privilégier l'utilisation de dispositifs réalisant les traitements de données en local à ceux réalisant un traitement distant.
- Privilégier l'utilisation de dispositifs équipés d'un bouton de désactivation physique du microphone.
- Privilégier des dispositifs permettant l'activation de l'écoute par pression manuelle sur le dispositif plutôt que par un mot-clé, ce qui vous donnera une plus grande maîtrise sur ses moments d'activations. À défaut, privilégier les dispositifs signalant via un signal sonore le début et la fin des périodes d'enregistrement et les activer lors de l'installation de l'assistant vocal.

Sur l'utilisation du dispositif

- Si vous ne souhaitez pas que des personnes écoutent vos conversations et que votre dispositif le permet, désactivez l'analyse de vos interactions pour les finalités d'amélioration du produit.
- Si vous ne souhaitez pas partager les données techniques, désactivez l'analyse de celles-ci pour les finalités d'amélioration du produit.
- Couper le micro/éteindre l'appareil lorsque vous ne souhaitez pas être écouté par l'assistant. À noter que certains dispositifs n'ont pas de bouton marche/arrêt et doivent donc être débranchés.
- Avertir les tiers (invités, personnel de maison, etc.) de l'enregistrement potentiel des conversations, ou couper le micro/éteindre l'appareil.
- Réciproquement, dans un lieu où vous êtes temporairement et dans lequel un assistant vocal est présent, demander au propriétaire de le désactiver ou le débrancher si vous ne souhaitez pas être enregistré.
- Lorsque qu'il est embarqué dans un dispositif dédié, positionner l'assistant vocal à un endroit où il sera bien en évidence et visible de tous.
- Vérifier régulièrement dans l'espace utilisateur l'historique des données enregistrées et supprimer les données confidentielles.

Nos conseils

- Être vigilant sur le fait que les propos tenus face à l'appareil peuvent enrichir votre profil publicitaire. La plupart des concepteurs d'assistants permettent d'afficher les segments publicitaires dans lesquels vous avez été catégorisés et de supprimer cette catégorisation.
- Privilégier des dispositifs ne nécessitant pas la création d'un compte utilisateur pour leur usage.
  - Lorsque le dispositif impose l'utilisation d'un compte utilisateur, ou lorsque certaines fonctionnalités rendent l'usage d'un compte nécessaire, évaluer s'il est préférable de lier un compte déjà existant ou, au contraire, de créer un compte dédié.
  - Lorsque l'usage d'un compte individuel est nécessaire pour certaines fonctionnalités, garder en mémoire que toute personne ayant accès à l'assistant sera en capacité de les utiliser une fois ce dernier installé, sauf si vous paramétrez des mesures d'authentification.
- Si l'assistant permet cette fonctionnalité, privilégier l'utilisation d'un mode « non connecté » (navigation privée) à ses comptes lorsque la connexion n'est pas nécessaire au traitement et à l'exécution de la commande passée.
- Ne connecter à l'assistant que des services qui présentent réellement une utilité, tout en considérant les risques pour la vie privée à partager des données intimes ou des fonctionnalités sensibles.
- Vérifier régulièrement quels sont les services connectés à l'assistant, et désactiver les services peu ou pas utilisés.
- Ne pas hésiter à contacter les services supports en cas de questions et à exercer ses droits auprès d'eux (par exemple le droit d'accès), et, le cas échéant, la CNIL.

Contrôler la monétisation de l'intime

Principalement destinés au domicile (ou à son prolongement qu'est le véhicule personnel) pour contrôler des objets connectés et des services de divertissement, les appareils dotés d'un assistant vocal se retrouvent au cœur de la vie du foyer. Dans de nombreux cas, les différentes interactions de l'utilisateur avec l'assistant alimentent un profil lié à ce dernier. Habitudes de vie (heure de lever et coucher), réglage du chauffage, goûts culturels, achats passés, centres d'intérêt, etc., toutes ces informations peuvent ensuite être utilisées à des fins de ciblage publicitaire.

Se souvenir de l'absence d'écran

Si la présence d'un écran compagnon est bien souvent nécessaire pour configurer son assistant, l'ambition des assistants vocaux est de proposer des interactions ne s'appuyant pas principalement sur un support visuel. Toutefois, sans écran, il est parfois difficile d'avoir un aperçu des traces enregistrées, de juger de la pertinence des suggestions qui sont faites, d'en savoir plus ou d'avoir accès à des réponses provenant d'autres sources.

Nos conseils

- Privilégier les dispositifs qui permettent la gestion des paramètres de l'appareil et l'effacement des données via l'interface vocale en plus de l'option via l'écran compagnon ou le compte utilisateur.
- Se rendre régulièrement sur le tableau de bord de gestion de l'assistant pour personnaliser ses fonctionnalités selon ses besoins. Par exemple, définir le moteur de recherche ou la source d'information utilisée par défaut.
- Ne pas hésiter à utiliser les fonctionnalités de l'assistant pour programmer des rappels des conseils présentés dans ce chapitre !

Encadrer les usages par les enfants

D'abord objet de curiosité, les assistants vocaux peuvent rapidement devenir une interface numérique particulièrement appréciée des enfants pour sa (relative) facilité de prise en main. S'il ne fait aucun doute qu'un ordinateur ou un smartphone ne doit pas être laissé dans les mains d'un jeune enfant sans supervision parentale, il essentiel de noter qu'il en va de même pour les interfaces pilotées uniquement par la voix.

Nos conseils

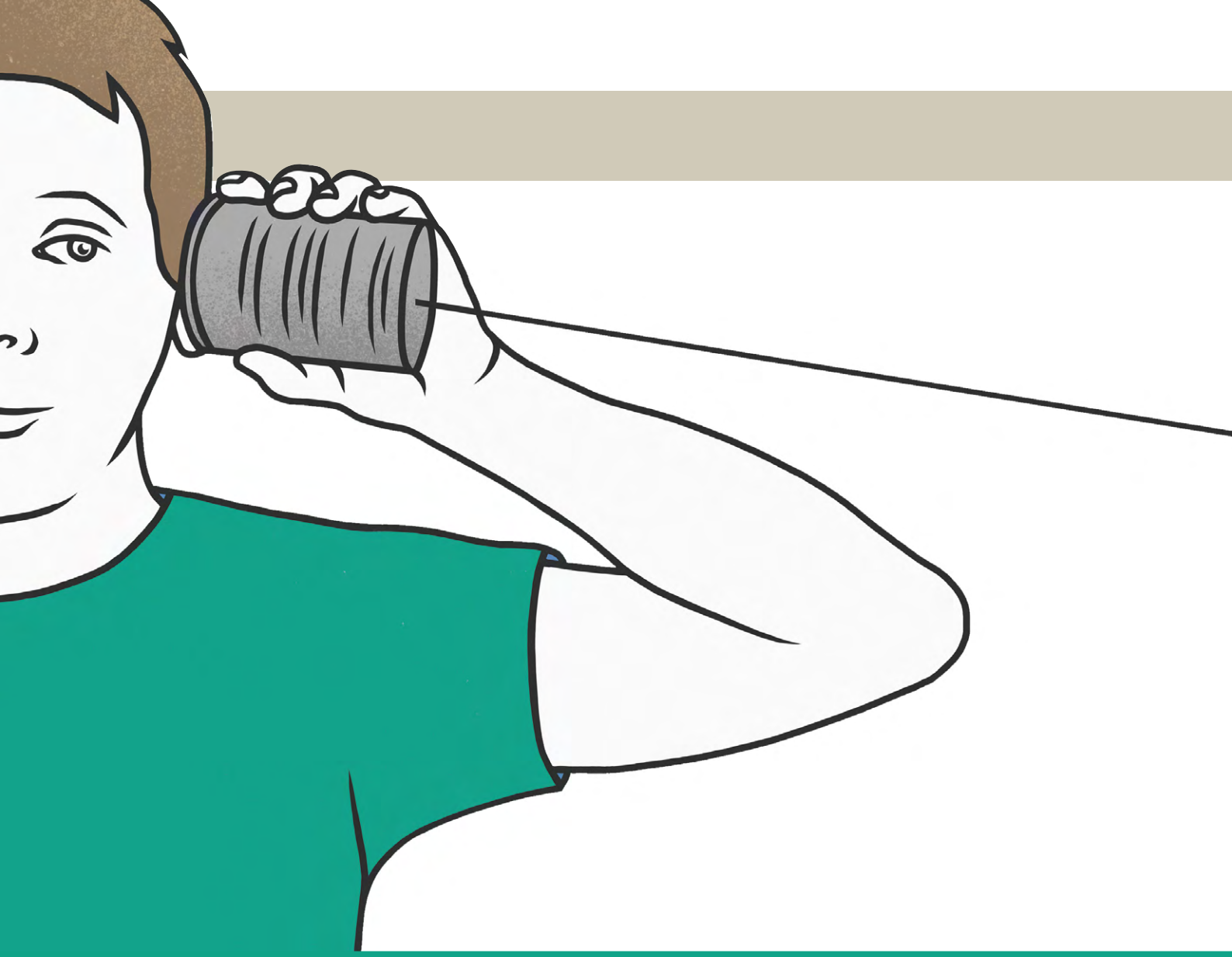
- Expliquer de façon pédagogique le mode de fonctionnement d'un assistant vocal et montrer les réglages simples (bouton de désactivation par exemple).
- Éviter de déployer ces dispositifs dans les espaces réservés aux enfants (chambre, salle de jeu, etc.).
- Encadrer les interactions des enfants avec le dispositif : rester dans la pièce lorsqu'ils l'utilisent, l'éteindre lorsqu'on n'est pas avec eux.
- Vérifier qu'il est bien réglé par défaut pour filtrer les informations à destination des enfants.
- Si un historique est enregistré, consulter les statistiques d'utilisation et, le cas échéant, les interactions passées, dans le respect de la vie privée de l'enfant.
- Supprimer de manière régulière cet historique.

Prévenir les risques de piratage

En fonction des choix paramétrés par l'utilisateur, différents services peuvent être accessibles par un assistant vocal. Toutefois, celui-ci n'offre pas toujours de possibilité d'authentification permettant de s'assurer de la légitimité de la personne passant la commande. Il convient donc de garder à l'esprit que l'assistant vocal peut, dans le cas où il est relié à de nombreux services (par exemple domotiques ou bancaires), être une brèche dans le système d'information du foyer.

Nos conseils

- Comme pour tout objet connecté, éviter les produits dont l'origine et le concepteur ne sont pas reconnus ou pour lesquels il n'est pas possible d'identifier facilement le responsable et un point de contact, idéalement en français.
- Choisir avec attention les services qui peuvent être pilotés par son assistant vocal et éviter ceux à risque (ouverture de porte, serrure, démarrage d'un véhicule, etc.).
- Faire attention à n'installer et n'accéder qu'à des applications légitimes, des pirates pouvant créer des applications malveillantes afin de collecter des données d'utilisateurs (numéro de compte ou de carte bancaire, mot de passe, adresse, contact, etc.).
- Paramétrer la sécurité du dispositif ou de certaines applications sensibles, à travers une authentification à deux facteurs (par exemple via une validation à effectuer suite à un envoi d'email ou SMS) si le dispositif le permet.
- Choisir avec soin l'activation dans l'assistant de services liés à ses comptes (mails, agenda, compte bancaire, appels, etc.) qui seraient susceptibles d'être accessibles par toute personne se trouvant dans la même pièce.
- Sécuriser le réseau (en particulier le Wi-Fi) auquel est connecté l'assistant.



Commission Nationale de l'Informatique et des Libertés
3 place de Fontenoy
TSA 80715
75334 PARIS CEDEX 07
Tél. +33 (0)1 53 73 22 22
www.cnil.fr

Septembre 2020

CNIL.
COMMISSION NATIONALE
INFORMATIQUE & LIBERTÉS