



Nicolas KUSZ

Université Paris 1 Panthéon-Sorbonne (PRISM Sorbonne)
nicolas.kusz@gmail.com

Jean-François LEMOINE

Université Paris 1 Panthéon-Sorbonne (PRISM Sorbonne)
ESSCA School of Management
jean-francois.lemoine@univ-paris1.fr

INTENTION D'ACHAT ET ASSISTANT VOCAL : COMMENT LE TYPE DE VOIX INFLUENCE LES REponses DES CONSOMMATEURS ?

Résumé : De nombreuses entreprises créent leur assistant vocal afin d'offrir un nouveau canal d'interaction à leurs consommateurs. Malgré les progrès majeurs des technologies de reconnaissance vocale et de dialogue homme-machine, la voix de synthèse appliquée à l'assistant vocal semble être une caractéristique n'ayant pas été suffisamment étudiée dans le cadre de travaux théoriques. Cette recherche propose d'étudier l'influence du type de voix de l'assistant vocal (humain versus artificiel) sur les réactions et les intentions comportementales des utilisateurs. Les résultats d'une expérimentation menée auprès de 435 répondants montrent qu'une voix humaine influence plus positivement l'anthropomorphisme perçu par les utilisateurs. En revanche, contrairement à la littérature portant sur les agents virtuels et les *chatbots*, notre étude révèle qu'une voix de synthèse plus artificielle permet de susciter de meilleures réactions sur la facilité d'utilisation perçue, l'utilité perçue et l'intention d'achat par le biais de l'assistant vocal.

Mots clefs : assistant vocal ; voix de synthèse ; anthropomorphisme ; utilité perçue ; facilité d'utilisation perçue.

PURCHASE INTENTION AND VOICE ASSISTANT: HOW DOES THE TYPE OF VOICE INFLUENCE CONSUMER REACTIONS?

Abstract : More and more companies are creating their voice assistant to offer a new channel of interaction to their consumers. Despite major progress in voice recognition and human-machine dialogue technologies, the synthetic voice applied to the voice assistant seems to be a characteristic that has not been sufficiently studied in the context of theoretical work. This research proposes to study the influence of the voice type of the voice assistant (human versus artificial) on the reactions and behavioral intentions of users. The results of an experiment carried out with 435 respondents show that a human voice has a more positive influence on the anthropomorphism perceived by users. On the other hand, contrary to the literature on virtual agents and chatbots, our study reveals that a more artificial synthetic voice can elicit better reactions on perceived ease of use, perceived usefulness and intention to purchase through the voice assistant.

Keywords : voice assistant ; synthetic voice ; anthropomorphism ; usefulness ; ease of use.

Résumé managérial

Chaque entreprise peut développer une application vocale et venir se brancher sur l'assistant vocal Google ou Alexa, et ainsi offrir un canal d'interaction supplémentaire à leurs clients en complément de leur site web ou de leur application mobile. Appelées « actions » pour Google et « skills » chez Amazon, ces applications vocales dites « tierces » ou « de marque » permettent de créer une discussion *via* une commande vocale telle que « Alexa, je veux parler à [nom de l'entreprise] », qui déclenche l'expérience vocale (avec des fonctionnalités et des services spécifiques) et le Dialogue Homme-Machine imaginés par chaque entreprise. Les entreprises ont aussi la possibilité d'opter pour une voix spécifique et unique pour leur assistant vocal plutôt que la voix du système Google ou Alexa, leur permettant de créer leur propre identité sonore. L'application vocale permet à l'utilisateur d'échanger directement avec une entreprise ou une marque comme il le ferait en venant naviguer sur un site web, de s'informer sur les produits ou les services, voire même acheter par la voix ou accomplir certaines actions spécifiques par commande vocale. Cette technologie ouvre ainsi de nouvelles perspectives inédites d'achat en ligne par le biais de la voix.

L'apport de notre recherche est double pour les managers : nous mettons en évidence le rôle majeur de la voix dans le cadre de la création d'un assistant vocal, puis nous renseignons les professionnels sur le type de voix à privilégier entre une voix humaine versus plus robotique. Nos résultats contribuent également à une meilleure compréhension de l'effet des caractéristiques anthropomorphiques de la voix d'un assistant vocal. Le type de voix induit un effet significatif sur la façon dont les utilisateurs souhaitent interagir avec l'assistant vocal, et comment ils réussissent finalement à le faire. La création d'une application vocale n'est pas une interface comme les autres telles qu'un chatbot ou un agent virtuel (et cela même s'ils sont dotés d'une voix de synthèse), l'absence d'écran apporte des réactions et des comportements utilisateurs spécifiques à ce type d'interface vocale. Tous les individus ne semblent pas à l'aise avec le fait de dialoguer avec un objet inerte telle qu'une enceinte connectée ; les utilisateurs ne comprennent pas exactement avec quoi et avec qui ils échangent, ils ne savent pas comment fonctionne une application vocale, ni le type de questions qu'ils peuvent poser. On observe finalement de meilleures réactions sur la facilité d'utilisation, l'utilité perçue et in-fine l'intention d'achat d'un produit ou d'un service dans le cas où l'assistant vocal présente une voix de synthèse qui, bien que dotée de caractéristiques anthropomorphiques, soit bien identifiée par les utilisateurs comme une voix artificielle, et non pas comme une voix humaine. Nos résultats mettent en lumière qu'une voix très humaine induit un dialogue beaucoup plus riche de la part des utilisateurs, qui se mettent alors à interagir avec l'assistant de la même manière qu'ils le font avec un humain, avec toute la richesse et la complexité du langage, oubliant les limites techniques de l'appareil. Nous proposons ainsi aux entreprises et aux concepteurs d'intégrer des caractéristiques artificielles dans la voix de l'assistant pour aider l'utilisateur à garder en mémoire qu'il interagit avec une machine et non pas avec un humain, et lui permettre d'interagir plus facilement avec l'appareil. L'ensemble de ces résultats peut ainsi aider les spécialistes du marketing à comprendre les effets générés par la voix de l'assistant sur les réactions de leurs clients, afin de produire des efforts marketing ciblés pour créer une identité sonore qui puisse apporter de meilleures performances pour leur assistant vocal.

Introduction

Les assistants vocaux se sont démocratisés dans un premier temps grâce aux enceintes connectées Google Home et Amazon Echo, mais désormais chaque entreprise a la possibilité de créer son propre assistant vocal ou de créer une application vocale (appelée « App Action » chez Google ou « Skill » chez Amazon) et la brancher sur le système vocal Google ou Amazon (McLean *et al.* 2021 ; Balakrishnan et Dwivedi, 2021 ; Vernuccio *et al.*, 2022). Fin 2023, il y avait 900 millions d'enceintes connectées dans le monde (15 millions en France), mais on comptait en réalité plus de 8,5 milliards d'assistants vocaux car les appareils prenant en charge la voix et pouvant accueillir un assistant vocal au même titre qu'une enceinte connectée sont de plus en plus nombreux : ordinateur, smartphone, télévision reliée à Internet (Smart TV), voiture connectée, montre connectée et beaucoup d'autres objets connectés (parfois même à l'insu des individus), assurant une diffusion extraordinairement rapide des assistants vocaux auprès du grand public (Statista, 2024). Contrairement aux *chatbots* et agents virtuels qui viennent en complément d'une navigation sur un site web ou une application mobile, les caractéristiques des assistants vocaux diffèrent des autres technologies ; ils peuvent s'utiliser (totalement ou presque) en autonomie sans aucune autre interface, ni écran, ni clavier (Natale, 2020). Loin d'être anodine, cette technologie remet totalement en perspective la manière dont les individus consomment du contenu sur le web, recherchent l'information et achètent des produits ou des services (McLean *et al.*, 2019 ; Ramadan *et al.*, 2019). Bien que la voix, universelle et innée, apparaisse comme l'interface la plus simple et naturelle pour échanger entre l'homme et la machine, elle fait naître aussi de nouveaux défis associés au design web ; sans visuel, ni aucun texte ou bouton pour guider ou susciter l'intérêt (Devillers *et al.*, 2020). Malgré plusieurs travaux menés sur les agents virtuels et des *chatbots* ayant mis en évidence l'influence positive d'une voix de synthèse anthropomorphe (qui imite l'homme) sur les réactions des utilisateurs, à notre connaissance aucune recherche n'a étudié les effets des caractéristiques vocales d'un assistant vocal sur les réactions des utilisateurs. Par conséquent, nous proposons de répondre à la problématique suivante : ***Dans quelle mesure le type de voix de l'assistant vocal influence-t-il les réactions et les intentions comportementales des consommateurs ?*** Pour y répondre, nous abordons dans une première partie la littérature sur la voix et ses effets sur les individus, puis nous présentons nos hypothèses de recherche. Nous détaillons dans la seconde partie l'expérimentation menée et présentons les résultats de l'étude. D'un point de vue théorique, cette recherche vise à enrichir la littérature sur les assistants vocaux et les effets de la voix. D'un point de vue managérial, cette recherche a pour ambition de proposer des pistes de réflexion pour aider les professionnels à identifier le type de voix de leur assistant vocal afin de répondre à leur objectif en matière de facilité d'utilisation, d'utilité perçue et d'intention d'achat en vocal.

1. Le cadre conceptuel de la recherche

S'intéresser aux effets de la voix d'un assistant vocal nous amène tout naturellement à étudier la littérature sur la voix de synthèse puis à ses effets sur les réactions des individus.

1.1. La voix de synthèse (d'un assistant vocal)

Concevoir des machines performantes sur le plan technique, capables de comprendre et parler en langage naturel, ne suffit plus pour en assurer leur adoption par les utilisateurs. Doter une machine de caractéristiques anthropomorphiques, telles un corps, un visage, une voix ou des attitudes humaines, tend à améliorer les réactions des individus (Guthrie, 1997). La voix est le moyen le plus efficace pour échanger entre les individus (Cornut, 2009), Mehrabian indiquait dès 1967 que seulement 7% du langage passe par les mots tandis que 38% de la communication passe en réalité par la voix. La voix de synthèse correspond à la production d'une voix artificielle qui imite la voix humaine (Natale, 2020). Ainsi, de nombreuses caractéristiques

vocales (Cornut, 2009) permettent d’humaniser les voix de synthèse des assistants vocaux ; le type de voix, le genre de la voix, la tonalité grave ou aiguë, le timbre de voix ou encore la hauteur, voici autant d’éléments qui apportent une chaleur humaine aux assistants (Dubiel *et al.*, 2018 ; Reinkemeier *et al.*, 2022). Enfin, les caractéristiques du dialogue homme-machine (DHM), telles que les hésitations, voire les silences tels des moments de réflexion ou de respiration sont autant de marqueurs qui façonnent la personnalité humaine d’un assistant vocal (Moriuchi, 2019 ; Rzepka *et al.*, 2020 ; Patrizi *et al.*, 2021 ; Balakrishnan et Dwivedi, 2021). Selon Natale (2020), la technologie de synthèse vocale et ses caractéristiques paralinguistiques ne permettraient plus aux individus de discerner une voix de synthèse d’une voix humaine tant elle semble réaliste. Selon Berriche et Benavent (2021), une voix humaine permet également de dissimuler les limites techniques des assistants vocaux.

1.2. Les effets de la voix de synthèse sur les réactions des utilisateurs

La voix remet totalement en question la manière dont les individus interagissent avec un système tel qu’un assistant vocal (McLean *et al.*, 2019 ; Ramadan *et al.*, 2019). La voix de l’assistant vocal change la donne : l’interface est « invisible », dépourvue de visuels, sans bouton sur lequel cliquer ni aucun texte pour susciter l’intérêt ou guider l’utilisateur (Velkovska et Zouinar, 2018). Naturellement, l’utilisateur a tendance à s’adresser à un assistant vocal de la même manière qu’avec un autre individu (Sundar *et al.*, 2017 ; Lopatovska, 2019 ; Canziani et MacSween, 2021). Bien qu’il n’existe pas, à notre connaissance, d’étude spécifique sur l’influence de la voix des assistants vocaux sur les réactions des consommateurs, des recherches antérieures menées sur les effets de la voix de synthèse suggèrent qu’une voix humaine appliquée à un agent virtuel ou un *chatbot* peut influencer positivement l’anthropomorphisme (Chérif et Lemoine, 2019). Toutefois, de précédents travaux révèlent que la maximisation du réalisme peut rendre un système repoussant (théorie de la vallée de l’étrange ; Mori, 1970). Pour Lopatovska *et al.* (2018) avec une voix très humaine les individus auraient aussi tendance à discuter avec un assistant vocal de la même manière qu’ils le feraient avec un humain en oubliant les limites techniques de l’appareil. Selon Abdulrahman et Richards (2022) les machines devraient avoir une voix identifiable, qui ne soit ni trop robotique ni trop humaine, pour générer de meilleures réactions. Une étude récente confirme également que le caractère artificiel d’une voix pour une IA offre parfois une meilleure cohérence et un confort plus grand pour les utilisateurs (Wang *et al.*, 2024). En outre, pour comprendre les facteurs qui influencent les intentions comportementales des individus envers les assistants vocaux, les chercheurs s’appuient sur les fondements théoriques de l’acceptation d’une technologie (Moriuchi, 2019 ; McLean *et al.*, 2019). Le modèle TAM (Davis, 1989) révèle notamment que plus l’utilité perçue et la facilité d’utilisation perçue sont élevées, plus l’attitude de l’individu envers le système est positive. L’acceptation semble encore balbutiante et des freins demeurent, les attentes élevées des utilisateurs pensant pouvoir dialoguer de façon naturelle font face à des performances encore limitées et à des frustrations majeures (Poirier, 2017 ; Velkovska *et al.*, 2018).

Au regard de cette littérature, nous nous attendons à ce que le type de voix de l’assistant vocal influence l’anthropomorphisme perçu. Nous formulons les hypothèses suivantes :

- H1a : Une voix de synthèse humaine influence plus positivement l’anthropomorphisme perçu de l’assistant vocal, qu’une voix de synthèse artificielle.

En revanche, les travaux de Abdulrahman et Richards (2022) nous invitent à formuler les hypothèses suivantes sur les effets du type de voix sur les réactions cognitives :

- H2a : Une voix de synthèse dotée de caractéristiques artificielles a un effet plus positif sur l’utilité perçue de l’assistant vocal, qu’une voix de synthèse très réaliste.
- H3a : Une voix de synthèse dotée de caractéristiques artificielles a un effet plus positif sur la facilité d’utilisation perçue, qu’une voix de synthèse très réaliste.

En complément, en nous appuyant sur le modèle TAM comme cadre d'analyse, nous supposons le rôle déterminant de l'utilité perçue et de la facilité d'utilisation perçue dans l'explication des intentions comportementales des individus (McLean *et al.*, 2019). Bien que les composantes du TAM évaluent l'intention d'utilisation d'un système (ici, l'assistant vocal), plusieurs travaux en marketing digital (McLean *et al.*, 2019 ; Balakrishnan et Dwivedi, 2021 ; Mari *et al.*, 2024) démontrent que l'évaluation positive d'un système peut se transférer à des comportements spécifiques générés par ce système (ex. la recherche, la commande ou l'achat d'un produit). Nous considérons ainsi que l'intention d'achat *via* l'assistant vocal est une intention comportementale qui découle directement de l'interaction avec l'assistant. Nous formulons les hypothèses des effets médiateurs de l'anthropomorphisme, de l'utilité perçue et de la facilité d'utilisation perçue sur la relation entre le type de voix et l'intention d'achat d'un produit ou service *via* l'assistant vocal, respectivement les hypothèses H1b, H2b et H3b. De plus, dans la lignée des recherches antérieures, il nous apparaît utile de nous interroger également sur le rôle médiateur en série de la facilité d'utilisation puis de l'utilité perçue (H3c). L'ensemble des hypothèses formulées sont présentées en figure 1.

2. Méthodologie – procédure expérimentale

Afin de tester la validité de nos hypothèses, une expérimentation a été conduite auprès de 435 utilisateurs (détail en annexe 1), pour cela nous avons créé un assistant vocal. La voix de l'assistant a été manipulée selon deux modalités : voix de synthèse artificielle ou voix de synthèse humaine, exposée aléatoirement à deux groupes de répondants.

Assistant vocal de l'expérimentation : Dans le cadre de notre étude, nous nous sommes concentrés sur la création d'un assistant vocal accessible depuis un ordinateur. D'un point de vue technique, notre assistant vocal était capable de comprendre une question formulée à l'oral et de répondre vocalement (iso Google, Alexa ou ChatGPT). Pour créer l'assistant vocal, nous nous sommes appuyés sur le logiciel Voiceflow (annexe 2) et sur l'ensemble des technologies offertes (reconnaissance vocale, traitement automatique du langage naturel, dialogue homme-machine, synthèse vocale). Nous avons opté pour deux synthèses vocales (annexe 3) : (i) une voix de synthèse très réaliste et humaine, puis (ii) une voix de synthèse plus artificielle.

Pré-test : La vérification de la réussite de la manipulation (manipulation check) entre les deux types de voix a été pré-testée auprès de 80 répondants. Les voix de synthèse artificielles pouvant être malgré tout jugées réalistes car ayant des attributs anthropomorphes, un test de comparaison de moyennes a été effectué à l'aide d'une échelle bipolaire « voix artificielle ↔ voix humaine ». L'analyse de variance indique une différence entre les deux conditions expérimentales ($M_{\text{voix humaine}} = 4,69$ et $M_{\text{voix artificielle}} = 3,21$), avec un résultat statistiquement significatif ($t = -3,98$; ddl = 78 ; $p < 0,001$). Les deux voix sont bien perçues comme différentes par les deux groupes de répondants. Nous avons veillé à maintenir les autres variables non manipulées (vitesse, tonalité, hauteur, timbre, prosodie) constantes (Malhotra *et al.*, 2011).

Services proposés : Comme le précisent Shih et Venkatesh (2004), la pertinence du choix de service proposé (pensé pour un usage vocal) et la récurrence d'usage (l'utilisateur ne pense pas spontanément à utiliser une technologie qu'il ne sollicite pas régulièrement) sont des facteurs clés de succès d'un assistant vocal. Nous avons ainsi créé un assistant vocal dans le domaine de la santé ; l'utilisateur avait le choix entre bénéficier d'une information santé ou bien-être de type « le saviez-vous ? » (ex. les sodas « zéro » peuvent faire grossir si on en consomme trop, bien qu'ils valent 0 calorie), ou découvrir le « produit du moment » (ex. un savon artisanal bio). Nous avons opté pour ce type de services et de produits puisqu'ils touchent un grand nombre d'individus, quel que soit leur âge, leur genre ou leur classe sociale (le secteur hygiène-beauté est l'un des secteurs les plus importants en e-commerce, selon la Fevad 2024).

Echelles de mesure : Nous avons utilisé des échelles de mesure déjà mobilisées spécifiquement dans le cadre d'une interaction avec un assistant vocal (annexe 4). Inspirées du modèle TAM, nous avons utilisé les échelles de Pitardi et Marriott (2021) pour l'utilité perçue et la facilité d'utilisation perçue. Afin de mesurer l'anthropomorphisme perçu, nous nous sommes appuyés sur l'échelle de Patrizi *et al.* (2021). Enfin pour mesurer l'intention d'achat *via* l'assistant vocal nous avons fait le choix de nous appuyer sur l'échelle de Balakrishnan et Dwivedi (2021). Tous les items ont été mesurés par des échelles de Likert en 7 échelons.

Collecte des données : Afin de diffuser notre questionnaire et de collecter nos données auprès des individus, nous avons eu recours au panéliste Panelabs. Le panel de consommateurs a permis d'assurer plus facilement nos critères de ciblage tout en respectant les critères démographiques de la population moyenne des internautes (genre, âge et catégorie socioprofessionnelle). Le questionnaire fut composé de trois parties ; (i) des questions générales sur l'utilisation des assistants vocaux, (ii) la mise en situation in-situ (les répondants pouvaient utiliser l'assistant vocal chez eux), puis (iii) des questions sur les perceptions des individus.

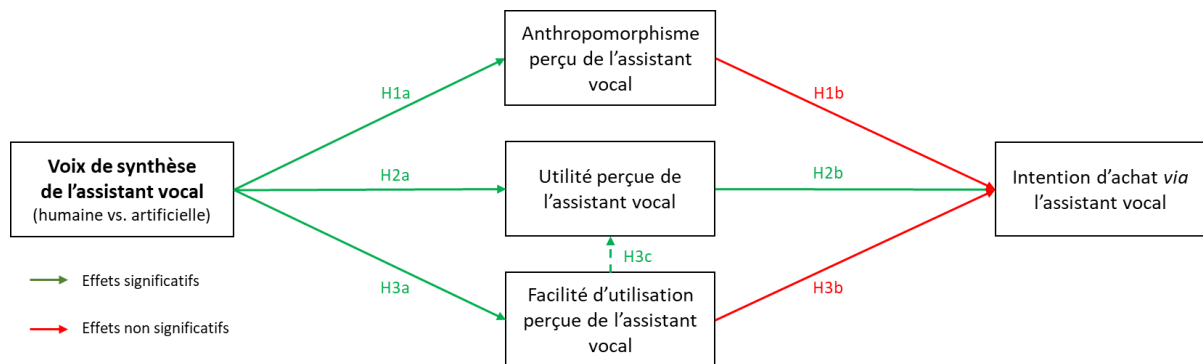
3. Résultats de l'étude

Afin de vérifier la validité de notre expérimentation, au préalable nous avons vérifié la validité et la fiabilité de nos échelles de mesure en réalisant une analyse factorielle exploratoire puis une analyse factorielle confirmatoire à l'aide des logiciels SPSS et AMOS (annexe 5). Enfin, avant de procéder à l'analyse des résultats nous avons également vérifié la validité convergente et discriminante de nos échelles de mesure (annexe 6).

Effets directs : Dans le but de tester la validité des hypothèses relatives à l'influence du type de voix de l'assistant vocal sur les réactions, nous avons réalisé des analyses de variance (ANOVA, détail en annexe 7). Les résultats indiquent un effet statistiquement significatif du type de voix sur l'anthropomorphisme perçu ($M_{\text{voix humaine}} = 4,838$ et $M_{\text{voix artificielle}} = 3,974$; $F = 37,624$; $p < 0,001$). Une voix humaine a une influence positive sur l'anthropomorphisme de l'assistant vocal, l'hypothèse H1a est validée. En complément, les résultats indiquent également un effet statistiquement significatif du type de voix sur l'utilité perçue ($M_{\text{voix humaine}} = 4,393$ et $M_{\text{voix artificielle}} = 4,684$; $F = 4,452$; $p = 0,035$) ainsi que sur la facilité d'utilisation perçue ($M_{\text{voix humaine}} = 5,226$ et $M_{\text{voix artificielle}} = 5,459$; $F = 4,537$; $p = 0,034$), les hypothèses H2a et H3a sont validées. En revanche, le sens de l'effet du type de voix s'oppose aux résultats observés sur l'anthropomorphisme perçu : une voix artificielle semble avoir un effet plus positif sur la facilité perçue et l'utilité perçue.

Effets médiateurs : Afin de tester nos hypothèses de médiation en parallèle, nous avons utilisé le modèle 4 de la macro-process de Hayes. Seule l'utilité perçue de l'assistant vocal (Coeff = $-,2338$; Intervalle de Confiance = $[-,4566 ; -,0208]$) joue un rôle médiateur dans la relation entre le type de voix et l'intention d'achat d'un produit ou d'un service par le biais de l'assistant vocal. En outre, l'effet direct du type de voix sur l'intention d'achat *via* l'assistant vocal n'est pas significatif ($p = 0,5591$), ce qui révèle une médiation totale. L'Hypothèse H2b est validée. En revanche, aucun effet significatif n'a été relevé concernant la médiation de la facilité d'utilisation de l'assistant vocal (Coeff = $-,0116$; IC = $[-,0410 ; ,0157]$), ou de l'anthropomorphisme de l'assistant vocal (Coeff = $,0508$; IC = $[-,0201 ; ,1255]$), les hypothèses H1b et H3b sont rejetées. En complément des effets de médiation en parallèle, bien que la facilité d'utilisation ne semble pas jouer un rôle médiateur, nos résultats indiquent malgré tout un effet de médiation en série : le type de voix de l'assistant vocal a un effet sur la facilité d'utilisation perçue qui à son tour influence l'utilité perçue et enfin l'intention d'achat *via* l'assistant vocal. Notre hypothèse de médiation en série H3c, testée avec le modèle 6 de Hayes, est ainsi validée ($a \times b \times c = -,1455$; IC = $[-,2860 ; -,0097]$).

Figure 1 : modèle de la recherche et principaux résultats



4. Apports de la recherche et conclusion

Sur le plan théorique, cette recherche présente trois apports principaux. Premièrement elle enrichit les travaux en marketing digital portant sur les effets de la voix de synthèse. En outre, si plusieurs recherches ont été menées sur les assistants vocaux aucune étude ne s'est intéressée aux effets de leur voix de synthèse sur les individus. Deuxièmement, nous mettons en évidence l'existence de liens directs entre la voix de l'assistant vocal et la facilité d'utilisation perçue, l'utilité perçue et l'anthropomorphisme perçu. Notre recherche révèle également l'existence d'un effet médiateur de l'utilité perçue entre le type de voix et les intentions comportementales. Enfin, nous mettons en évidence un effet de médiation en série de la facilité d'utilisation perçue puis de l'utilité perçue entre le type de voix de l'assistant vocal et l'intention d'achat d'un produit ou d'un service par le biais de ce dernier. Troisièmement, notre recherche s'oppose à la littérature menée sur les agents virtuels et les *chatbots*, bien qu'une voix humaine semble avoir un effet significatif et positif sur l'anthropomorphisme perçu, nos résultats suggèrent que l'anthropomorphisme ne suffit pas à déclencher un achat *via* un assistant vocal. L'inclusion de cette variable dans notre modèle permet de montrer que l'effet d'une voix humaine est limité à des perceptions sociales, mais qu'il ne se prolonge pas automatiquement sur les intentions comportementales. Nous mettons en évidence qu'un assistant vocal doté d'une voix de synthèse plus artificielle semble avoir une influence plus positive sur sa facilité d'utilisation, l'utilité perçue et l'intention comportementale des utilisateurs. Ce résultat, *a priori* contre-intuitif, est cohérent avec certaines études récentes sur l'interaction homme-machine (Abdulrahman & Richards, 2022 ; Wang *et al.*, 2024) et avec la théorie de la vallée de l'étrange. L'assistant vocal ne disposant pas de représentation visuelle, sans visage ni corps permettant à l'individu de comprendre qu'il échange avec une machine comme un agent virtuel, une voix trop humaine (qui tente de se faire passer pour un humain) semble affecter négativement les réactions des individus. Une voix humaine semble rendre à tort l'assistant vocal plus performant qu'il ne l'est réellement, l'utilisateur semble alors oublier qu'il s'adresse à une machine.

Au niveau managérial, nous recommandons aux professionnels de privilégier une voix de synthèse artificielle identifiable comme telle et qui permette à l'utilisateur de garder en mémoire qu'il échange avec une machine et non avec un humain, pour ajuster ses requêtes en conséquence. La voix offre de formidables opportunités pour créer une expérience en ligne propre à chaque entreprise et se différencier (Reeves *et al.*, 2019).

Notre recherche comporte malgré tout plusieurs limites qu'il convient de souligner. Elle présente une validité externe réduite compte tenu des particularités du terminal utilisé et de la catégorie de produits retenue. Par ailleurs, il serait intéressant d'étudier les effets de la voix sur l'intention d'usage de l'assistant vocal. Enfin, nous n'avons pas tenu compte des autres caractéristiques de la voix de synthèse (fréquence, intensité, vitesse, hauteur, timbre ou genre).

Références

- Abdulrahman, A., Richards, D. (2022). Is natural necessary? Human voice versus synthetic voice for intelligent virtual agents. *Multimodal Technologies and Interaction*, 6, 51.
- Balakrishnan, J., Dwivedi, Y. K. (2021). Conversational commerce: Entering the next stage of AI-powered digital assistants. *Annals of Operations Research*.
- Berriche, A., Benavent, C. (2021). Comprendre les profils des voicenautes et leur intention d'achat par les assistants vocaux : Les apports de la théorie du conflit décisionnel. 20th International Marketing Trends Conference, Digital Conference.
- Canziani, B., MacSween, S. (2021). Consumer acceptance of voice-activated smart home devices for product information seeking and online ordering. *Computers in Human Behavior*, 106714.
- Chérif, E., Lemoine, J.F. (2019). Les conseillers virtuels anthropomorphes et les réactions des internautes : une expérimentation portant sur la voix du conseiller. *Recherche et Applications en Marketing (French Edition)*, 34(1), 29-49.
- Cornut, G. (2009). *La voix*. Paris : Presses Universitaires de France (« Que sais-je ? »).
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13, 319-340.
- Devillers, L., Masure, A., Obin, N. (2020). Assistants connectés : des ordres dans la voix. Interview dans le cadre de l'émission Magnétique, Radio Télévision Suisse, Lausanne.
- Dubiel, M., Halvey, M., Azzopardi, L. (2018). A survey investigating usage of virtual personal assistants. *Computer Science ArXiv*, abs/1807.04606.
- Guthrie, S. E. (1997). Anthropomorphism: A definition and a theory. In R. Mitchell, N. Thompson, & H. Miles (Eds.), *Anthropomorphism, anecdotes, and animals* (pp. 50-58). Albany : State University of New York Press.
- Lopatovska, I. (2019). Overview of the intelligent personal assistants. *Ukrainian Journal on Library and Information Science*, 72-79.
- Lopatovska, I., Rink, K., Knight, I., Raines, K., Cosenza, K., Williams, H., Sorsche, P., Hirsch, D., Li, Q., Martinez, A. (2018). Talk to me: Exploring user interactions with the Amazon Alexa. *Journal of Librarianship and Information Science*.
- Malhotra, N., Décaudin, J.-M., Bougerra, A., Bories, D. (2011). *Études marketing + logiciel SPSS* (6^e éd.). Pearson Education France.
- Mari, A., Mandelli, A., Algesheimer, R. (2024). Empathic voice assistants: Enhancing consumer responses in voice commerce. *Journal of Business Research*, 175.
- McLean, G., Osei-Frimpong, K. (2019). Hey Alexa... Examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*, 99, 28-37.
- McLean, G., Osei-Frimpong, K., Barhorst, J. (2021). Alexa, do voice assistants influence consumer brand engagement? – Examining the role of AI powered voice assistants in influencing consumer brand engagement. *Journal of Business Research*, 124, 312-328.

- Mehrabian, A. (1967). Orientation behaviors and nonverbal attitude communication. *Journal of Communication*, 16, 324-332.
- Mori, M. (1970). Bukimi no tani genshō (Le phénomène de la vallée de l'étrange). *Energy*, 7(4), 33-35.
- Moriuchi, E. (2019). Okay, Google! An empirical study on voice assistants on consumer engagement and loyalty. *Psychology & Marketing*.
- Natale, S. (2020). To believe in Siri: A critical analysis of AI voice assistants. *Communicative Figurations Working Papers*, 32, 130-146.
- Patrizi, M., Vernuccio, M., Pastore, A. (2021). Developing voice-based branding: Insights from the Mercedes case. *Journal of Product & Brand Management*, 30(5), 726-739.
- Pitardi, V., Marriott, H. R. (2021). Alexa, she's not human but... Unveiling the drivers of consumers' trust in voice-based artificial intelligence. *Psychology & Marketing*.
- Poirier, F. (2017). Quelle modalité pour l'interaction avec les petits appareils mobiles et vestimentaires : texte ou vocal ? Comment choisir entre clavier et assistant personnel ? Dans 29^e conférence francophone sur l'Interaction Homme-Machine (pp. A-67). Poitiers, France. [Disponible sur HAL : hal-01577674]
- Ramadan, Z., Farah, M., Audi, H. (2019). The advent of the voice moment of truth: The case of Amazon's Alexa. *International Conference on Advances in National Brand*.
- Reeves, S., Porcheron, M., Fischer, J. E., Candello, H., McMillan, D., McGregor, M., Moore, R. J., Sikveland, R., Taylor, A. S., Velkovska, J., Zouinar, M. (2018). Voice-based conversational UX studies and design. Dans *CHI Conference on Human Factors in Computing Systems* (article W38, pp. 1-8). ACM.
- Reinkemeier, F., Gnewuch, U., Toporowski, W. (2022). Can humanizing voice assistants unleash the potential of voice commerce? Dans *International Conference on Information Systems*.
- Rzepka, C., Berger, B., Hess, T. (2020). Why another customer channel? Consumers' perceived benefits and costs of voice commerce. Dans *53rd Hawaii International Conference on System Sciences*.
- Shih, E., Venkatesh, A. (2004). Beyond adoption: Development and application of a use-diffusion model. *Journal of Marketing*, 68.
- Sundar, S. S., Jung, E. H., Waddell, F. T., Kim, K. J. (2017). Cheery companions or serious assistants? Role and demeanour congruity as predictors of robot attraction and use intentions among senior citizens. *International Journal of Human-Computer Studies*, 97, 88-97.
- Velkovska, J., Zouinar, M. (2018). The illusion of natural conversation: Interacting with smart assistants in home settings. Dans *2018 CHI Conference on Human Factors in Computing Systems*. ACM.
- Vernuccio, M., Patrizi, M., Seric, M., Pastore, A. (2022). The perceptual antecedents of brand anthropomorphism in the name-brand voice assistant context. *Journal of Brand Management*.
- Wang, X., Zhang, Z., Jiang, Q. (2024). The effectiveness of human vs. AI voice-over in short video advertisements: A cognitive load theory perspective. *Journal of Retailing and Consumer Services*, 81, 104005.

Annexes

Annexe 1 : Echantillon final

Tableau 18 : Caractéristiques de l'échantillon final et des deux sous-échantillons

Caractéristiques de l'échantillon final		N total (%)	N voix 1 (%)	N voix 2 (%)
Sexe ($\chi^2 = 1,60$; ddl = 2 ; p = ,450)	Femme	230 (53%)	113 (55%)	117 (51%)
	Homme	205 (47%)	92 (45%)	113 (49%)
Age ($\chi^2 = 3,23$; ddl = 3 ; p = ,357)	18-24 ans	26 (6%)	11 (5%)	15 (7%)
	25-39 ans	148 (34%)	64 (31%)	84 (37%)
	40-59 ans	191 (44%)	91 (44%)	100 (43%)
	60 ans et +	70 (16%)	39 (20%)	31 (13%)
Niveau d'études ($\chi^2 = 5,51$; ddl = 4 ; p = ,239)	BEP - CAP	55 (13%)	30 (15%)	23 (11%)
	Baccalauréat	87 (20%)	40 (20%)	47 (20%)
	Bac+2	98 (23%)	53 (26%)	45 (20%)
	Bac+ 3/4	88 (20%)	39 (19%)	49 (21%)
	Bac+5 et plus	107 (24%)	43 (20%)	66 (28%)
CSP ($\chi^2 = 4,93$; ddl = 3 ; p = ,177)	CSP+	142 (33%)	59 (29%)	83 (36%)
	CSP-	195 (45%)	92 (45%)	103 (45%)
	Retraité	49 (11%)	29 (14%)	20 (9%)
	Sans activité	49 (11%)	25 (12%)	24 (10%)
Possède un assistant vocal ($\chi^2 = 0,334$; ddl = 1 ; p = ,563)	Oui	263 (60%)	121 (59%)	142 (62%)
	Non	172 (40%)	84 (41%)	88 (38%)

Annexe 2 : Notre assistant vocal créé depuis Voiceflow

SELECT CHANNEL

What channel do you want to create for?

Choose a channel for your first Voiceflow project

Amazon Alexa
Design, prototype and publish Alexa Skills for Amazon Alexa.

DESIGN & PROTOTYPE PUBLISH

Google Assistant
Design, prototype and publish Google Actions for Google Assistant.

DESIGN & PROTOTYPE PUBLISH

Custom Assistant
Design, prototype and export for any conversational channel.

DESIGN & PROTOTYPE

IVR
Design, prototype and export a conversational IVR.

DESIGN & PROTOTYPE

Chatbot
Design, prototype and export a conversational chatbot.

DESIGN & PROTOTYPE

Mobile App
Design, prototype and export a mobile voice assistant.

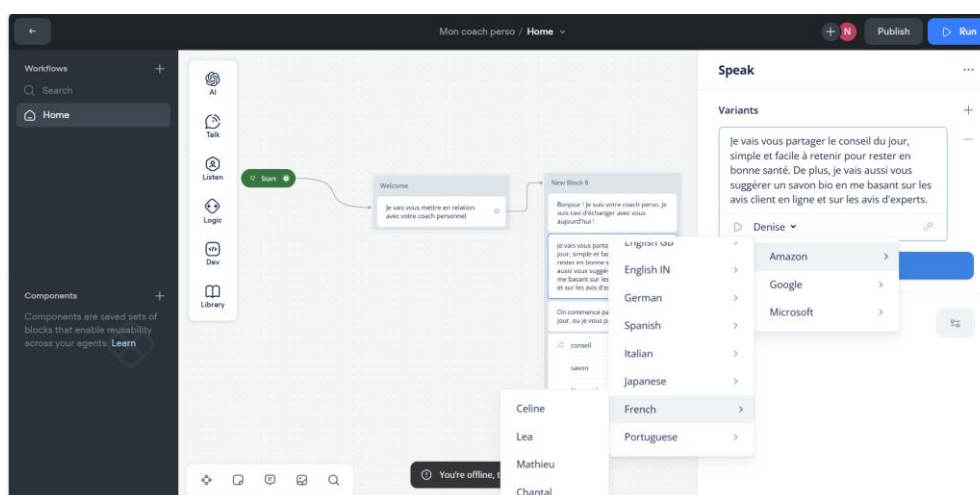
DESIGN & PROTOTYPE

Vous avez été invité à avoir une conversation avec **Mon coach perso**

Démarrer la conversation

8

Annexe 3 : Sélection des voix de synthèse pour notre assistant vocal



Annexe 4 : Synthèse des échelles de mesure pour le questionnaire final

Variables mesurées	Items	Codage	Sources
Utilité perçue de l'assistant vocal	L'assistant vocal fournit des informations de bonne qualité	PU1	Pitardi et Marriott (2021)
	Je gère plus efficacement mon temps lorsque je recueille des informations à l'aide de l'assistant vocal	PU2	
	Utiliser l'assistant vocal améliore ma performance pour rechercher une information	PU3	
	Dans l'ensemble, je trouve l'assistant vocal utile lorsque je recherche une information	PU4	
Facilité d'utilisation perçue de l'assistant vocal	Apprendre à utiliser l'assistant vocal est facile	PEU1	Pitardi et Marriott (2021)
	Mes interactions avec l'assistant vocal sont claires et compréhensibles	PEU2	
	Je trouve l'assistant vocal facile à utiliser	PEU3	
	Je trouve qu'il est facile de devenir habile pour interagir avec l'assistant vocal	PEU4	
Anthropomorphisme perçu de la voix de l'assistant	La voix de l'assistant vocal semble être humaine	HLV1	Patrizi <i>et al.</i> (2021)
	Le discours de l'assistant vocal semble être humain	HLV2	
	L'accent de l'assistant vocal semble être humain	HLV3	
	La qualité de la voix de l'assistant vocal semble être humaine	HLV4	
Intentions d'achat d'un produit ou service <i>via</i> l'assistant	Je souhaite acheter <i>via</i> l'assistant vocal	BUY1	Balakrishnan et Dwivedi (2021)
	Je souhaite renouveler mon expérience avec l'assistant vocal	BUY2	
	Je réaliserai mon prochain achat <i>via</i> l'assistant vocal	BUY3	
	Je suis enclin à acheter des produits ou des services <i>via</i> l'assistant vocal	BUY4	

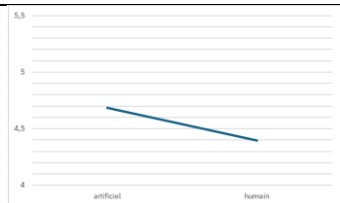
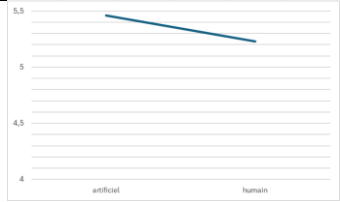
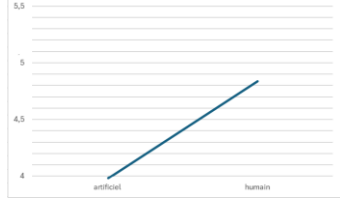
Annexe 5 : Synthèse des indices de validité et de fiabilité des échelles de mesure

Variables	Codage	Items	Alpha (α) de Cronbach >0,70	Rh $\hat{o}$ de Jöreskog >0,70	Validité convergente VME>0,5	Validité discriminante VME>corr. ²
Utilité perçue de l'assistant vocal	PU	4	0,936	0,938	0,791	confirmée
Facilité d'utilisation perçue de l'assistant vocal	PEU	4	0,920	0,923	0,750	confirmée
Anthropomorphisme perçu de l'assistant	HLV	4	0,945	0,946	0,815	confirmée
Intention d'achat par le biais de l'assistant	BUY	4	0,941	0,945	0,813	confirmée

Annexe 6 : Analyse de la validité discriminante des construits

Variables	VME	Carré des corrélations de chaque échelle avec les autres construits			
		PU	PEU	HLV	BUY
PU	0,791	0,791			
PEU	0,750	0,388	0,750		
HLV	0,815	0,214	0,140	0,815	
BUY	0,813	0,579	0,209	0,150	0,813

Annexe 7 : ANOVA et diagrammes de l'effet de type de voix sur les 3 réactions

Variable PU		M	σ	F	p	η^2	
Type de voix	artificiel	4,684	1,379	4,452	,035	,010	
	humain	4,393	1,481				
Test de Levene		F = 1,436 ; p = 0,231>5% (NS)					
Variable PEU		M	σ	F	p	η^2	
Type de voix	artificiel	5,459	1,040	4,537	,034	,010	
	humain	5,226	1,225				
Test de Levene		F = 1,614 ; p = 0,205>5% (NS)					
Variable HLV		M	σ	F	p	η^2	
Type de voix	artificiel	3,974	1,592	37,624	0,000	0,080	
	humain	4,838	1,343				
Test de Levene		F = 9,327 ; p = 0,002<5% (S)					
Test Kruskal Wallis		$\chi^2 = 11,516$; p = 0,001<5% (S)					