Voice app (Google Action and Alexa Skill) : the influence of the assistant's voice on users

Nicolas KUSZ
Université Paris 1 Panthéon-Sorbonne
Jean-François Lemoine
Université Paris 1 Panthéon-Sorbonne

Paper from the 52nd Annual EMAC Conference, Odense/Denmark, May 23-26, 2023

VOICE APP *(GOOGLE ACTIONS* AND *ALEXA SKILLS)* : THE INFLUENCE OF THE ASSISTANT'S VOICE ON TRUST AND SOCIAL PRESENCE

Abstract: Voice-activated virtual assistants (Google, Alexa, etc.) are becoming more and more a part of people's daily lives. Although their usefulness and functionalities are increasing, the voice interface between man and machine has not been sufficiently taken into account by companies that have developed their voice application (i.e. Alexa Skills or Google Actions) yet: the synthetic voice applied to the technology. Our research aims to measure the impact of the assistant's voice on consumers reactions. Based on 15 interviews of users, the results reveal that the type of voice of the assistant influences trust in the assistant and social presence. However, in contrast to studies conducted on virtual agents and chatbots, our research indicates that realism influences trust; a synthetic voice that sounds exactly like a human degrades the perception of the voice assistant.

Keywords: voice assistant; trust; social presence.

Track: Digital Marketing & Social Media.

1. Introduction

Made famous by the voice-activated connected speakers Google Home (Google Assistant) and Amazon Echo (Alexa) released respectively in France in 2017 and 2018, the virtual voice-activated assistant is defined as an intelligent system connected to the Internet, capable of understanding a question formulated orally by a human, then to perform certain actions or to give a response by synthetic voice (Lahoual and Fréjus, 2018). Assistants now offer many functionalities: information search, conversations with brands, online shopping, office or administrative tasks, entertainment or even home automation (Ammari et al., 2019; Gouliaéva et al., 2020). For their part, companies have the possibility of connecting a voice version of their website (i.e. a voice application) to virtual voice-activated assistants, to offer a new channel of interaction with their customers (Mclean et al. 2021). At the end of 2019, there were 200 million connected speakers worldwide (3 million in France), but there were actually more than 2.5 billion voice assistants (50 million in France) because there are more and more voice-activated devices, like the smartphones, speakers, smart TVs and many other connected objects, ensuring very quick adoption (Médiamétrie 2018, CSA 2019 and Hadopi 2019 studies). Unlike chatbots and virtual agents that complement browsing on a website or mobile application, the characteristics of voice assistants differ from other technologies; they can be used (totally or almost) independently without any other interface and, thus, replace classic web browsing with a screen and a keyboard (Natale, 2020). This technology completely puts into perspective the way individuals consume content on the web, search for information and buy products or services (Mclean et al., 2019; Ramadan et al., 2019). The voice, much more than the image, touches the imagination, invites pleasure and confidence. Although the voice, universal and innate, appears to be the most obvious solution for exchanging between man and machine, it also gives opportunities to new challenges associated with web design; with no visual reference or text to guide or spark interest, no buttons or links to click and navigate (Devillers, Masure & Obin, 2020). Given the recent arrival of virtual voice-activated assistants to the general public, this technology is still a new field of research and the work is mainly focused on understanding the first uses. The study problem that we have identified can be summarized as follows: does the voice of the assistant generated specific reactions ? The theoretical interest of this research is to highlight the effects of virtual assistant's voice on the reactions of users. From a managerial point of view, this research aims to propose lines of thought for brands wishing to develop their voice assistant.

2. The conceptual framework of the research

Being interested in the effects of assistant's voice on the user reactions naturally leads us to study the literature on human-machine interfaces and the voice. All of these two areas of research will provide a theoretical framework on the effects of the voice on users.

2.1 <u>Voice as Human-Machine Interface</u>

The success of voice assistants is also largely explained by the ease of use of its interface (Ammari et al., 2019; Rzepka et al., 2020), in this case the voice, as well as the emotion that is able to send back to its user. It easier to interact with systems thanks to IHMs, there was the keyboard and the screen, the mouse, then the touch screens and the touch, and finally the voice. The virtual voice-activated assistant changes the situation, since here the HMI is "invisible", without visuals, without design, without logo or without button to click (Velkovska and Zouinar, 2018). The voice as Man-Machine Dialogue (DHM) considerably influences behaviors, orally the user will express himself 3 times faster (Poirier, 2017). One can wonder if communicating with a system in the same way as we communicate with a person, would this not ultimately be the most successful HMI there is? The human voice is a mechanical process composed of very rich characteristics: frequency, intensity, pitch, timbre of the voice and finally prosody. This is obviously the most effective way to communicate between individuals (Cornut, 2009). The so-called synthetic voice, on the other hand, corresponds to the production of an artificial voice that imitates the human voice. Synthetic voices were first mechanical (1780), then electronic (1938), and finally computerized. If, before the 2000s, robotic synthetic voices were still quite far from a human voice and not very expressive, now they are modeled on the basis of real human voices, where actors recite many sentences in a loop allowing a machine to produce any sentence based on this voice. Anthropomorphism also humanizes the characteristics of synthetic voices; the sound of the voice obviously (robotic or human), but also onomatopoeia, hesitations, even silences such as moments of reflection or breathing, these are all markers that humanize the assistants (Dubiel et al., 2018). Finally, the characteristics of the assistant's voice, such as familiarity or formality, serious or friendly, low or high tone, gender, the ability to be humorous are all characteristics that shape "the identity voice" of a brand through its voice assistant. Not very expressive in the 2000s (Stern et al., 1999; Nass and Moon, 2000), now the assistants' voices perfectly imitate that of a real person (Shen et al., 2018). According to Natale (2020), voice synthesis technology and its paralinguistic characteristics (based on real human voices) would no longer allow individuals to discern a

synthesized voice from a human voice as mimicry is pushed to its climax. According to Obin and Roebel (2016), from now on "speech is no longer the privilege of humans".

2.2 <u>The effects of synthetic voice on users</u>

The effects of machine voice on user behavior are numerous. Mehrabian indicated as early as 1967 that only 7% of language passes through words, and 38% through the voice in reality. Even if at this stage, to our knowledge, there is no study on the assistant's voice, there are however many studies on the effects of the voice of virtual agents and chatbots on the behavioral intentions and reactions of consumers, on many dependent variables traditionally studied in marketing (Lee and Nass, 2003; Qiu and Benbasat, 2009; Chérif and Lemoine 2019). The user tends to interact naturally with the machine as with a human (Porcheron et al., 2018). Among the main effects of the voice, we can cite trust in technology, such as perceived honesty and credibility or the consideration of consumer expectations and motivations (Lemoine and Notebaert, 2011). In the field of voice assistants, early work also confirms the impact of the voice on perceived trust in technology (Pagani et al. 2019; Liao et al., 2020). The literature also states that the theory of social presence (i.e. the ability of a technology or interface to release a form of human warmth and allow the user to have the feeling of being in the presence of another individual) could also apply to voice assistants and motivate users to adopt the technology (Sundar et al, 2017; Chattaraman et al., 2019; Mclean et al., 2019 and 2021). The literature suggests that individuals interact with a voice assistant instinctively, in the same way as they would with a human and thus develop a special relationship with the device (e.g. users naturally tend to say hello to their assistant or the thank), we can deduce that social presence plays an essential part in user behavior (Kiseleva et al., 2016; Lopatovska et al., 2018). Finally, the work carried out on virtual agents also sheds light on the effects measured on other variables, such as attachment, realism or amusement.

3. Methodology

This qualitative study conducted with consumers is based on 15 semi-structured interviews, conducted face-to-face from may to october 2021. In order to benefit from a representative sample of users of this type of technology, we have selected a varied audience in terms of socio-demographic variables, but also proximity to the device (frequent or occasional use, seasoned users or not).

In a second step, we used the method of protocols, in order to circumvent any brakes on the perception of technology. For our study, we invited respondents to manipulate and interact with voice assistants. This allowed us to highlight the emotions felt by the individuals interviewed and to observe their behavior in

Variables	Characteristics	Number	Percentage
Sex	Man	8	53%
	Woman	7	47%
Age	-18 ans	2	13%
	18-25	2	13%
	26-35	3	20%
	36-50	5	34%
	+51	3	20%
Experienced user or new user	Experienced user	11	73%
	New user	4	27%
Frequency of use	Every day	11	73%
	Every week	3	20%
	Every month	1	7%

real situations of use. We asked the interviewees to discuss with five branded voice assistants, selected from a benchmark of around fifty assistants offered by certain brands.

4. Results and discussion

Our analysis reveals that the voice type of assistants has a significant impact on user reactions and emotions.

4.1 Perception of the voice HMI

The ease of use provided by the voice HMI stands out as one of the main success factors of voice assistants and thus confirms the literature (Kowalczuk, 2018), "there is still a story of ease. I can ask my son to ask for the weather." The HMI is a unanimous success with respondents, "I like my voice assistants because I don't need to type, you see... I just say OK Google and it answers me", or "There, without instructions , it's still very simple to use". The device also reduces the cognitive load of searching for information on a search engine and the web, "I'm going to go to Google or Yahoo and I'm going to have to look for information myself on lots of websites, whereas with a voice assistant I don't need to ask myself which website to get the information from, the assistant will do it automatically". Conversely, some consumers are afraid of falling into a certain form of harmful ease provided by this HMI, "even if it provides ease of answering questions, if it is fast, it is also a bit soothing, I would say . Soothing because it takes away the need or the will to get up and get the information yourself", some users add "I don't want my son to get used to this automaticity. […] to give him the impression that everything is easy and that the information can be found like that, with the snap of his fingers". Finally, despite obvious ease of use, respondents mentioned the blank page syndrome, "using a voice assistant is when you know what you can do with it […] and in fact, as you don't know, you don't don't use it", a disadvantage noted by many respondents "when you buy

something you don't even have a user manual, so you don't really know how to really use them 100%. What are good keywords? ".

4.2 Voice and its effects on the consumer

The voice emerges as one of the determining variables that motivate the use of a voice assistant, one respondent indicates "finally the voice of the assistants was not something that I paid too much attention to until now, but with the tests I realize unconsciously that the voice matters a lot". Referring to synthetic voices, one respondent said "I really don't feel like it's a machine, it's almost scary. The tone is also very impressive, the non-verbal language, and it is extremely natural".

For our study, we selected branded voice applications (backed by Google or Alexa systems) with a wide variety of voice characteristics, among which we can cite: gender, speaking speed, tone, pitch, timbre, prosody and type of voice (human or robotic).

Google Actions or Skills Alexa	Features	Voice's sex	Type of synthetic voice	Tone	Familiarity or Formal
Guerlain	Online fragrance design	woman	artificial	serious	Formal
Hello Bank	Account Access	woman	realistic	Friendly playful	Formal
Boulanger	Find a store, product information	man	artificial	serious	Formal
Oui.scnf	Find a trip	woman	artificial	serious	Familiarity or Formal
BNP Paribas	Account Access	man	realistic	Friendly playful	Formal

After interacting with branded voice apps, our study found that assistants with a more robotic synthetic voice were more successful, "OuiSCNF's voice I found more human, but not 100% either. I found it more pleasant", or even speaking of another robotic synthetic voice "the voice I didn't give myself any particular thought […] there was a little robotic side to it but that didn't bother me ". Contrary to the work carried out on chatbots and virtual agents (Lee and Nass, 2003; Wang et al., 2007; Chérif and Lemoine 2019), a human voice that nevertheless has some robotic characteristics thus seems to generate a positive effect on trust, "I don't really like human voices, I prefer when it's more robotic, a bit spatial because it's more modern, it's more futuristic...it's a machine, I expect a voice machine", another respondent adds "it doesn't bother me that the voice is robotic, it doesn't bother me on the contrary, that way I'm aware that I'm exchanging with a machine and not a human, I don't want the robot to imitate the human, it's not a human". The perceived realism strongly influences the reactions of users, "if it was a human voice it would be weirder than if it was a robot voice [...] I prefer this robot side because a human voice would put me more uncomfortable. A human voice tends to humanize technology even more and this boundary between human and machine could scare me", a phenomenon which is reminiscent of the valley of the uncanny (MacDorman, 2019), "you imagine I'm talking to someone thinking it's a human and finally I realize it was a robot…

frankly I wouldn't trust it". The voice assistant having no visual representation unlike virtual agents and chatbots, a respondent adds that with this type of technology it is necessary to have "a few clues that allow to clearly identify the artificial character of the voice and the system with whom I speak". An assistant with a very realistic human voice seems disturbing, "I find the voice more unhealthy, it looks more stupid" or even "this does not inspire confidence in me". It even brings a certain form of mistrust, "I don't really know what I have to do, he has a human voice but it's weird, I will trust him less". In addition, the voice seems to have an impact on social presence, as some respondents indicate, "sometimes it gives a nice side to talking to someone", or even "I preferred the voice of Google Home because we have the impression that he was no longer there with us". One respondent told us "it also brings a presence, even if it is not human, it is still more pleasant than staying behind a screen to type, it is more lively, more playful and more pleasant". The anthropomorphic character adds a little more to the technology, "what's nice is that the assistant has a first name, so it allows you to get closer to him and the voice". Conversely, when users interact with a voice assistant benefiting from a very realistic human voice, they seem to be in resistance "I had the impression that it was an answering machine completely".

5. Contribution of research and conclusion

The theoretical interest of this research is to study the multidisciplinary concepts of HMIs and voice, making it possible to highlight the effects of the assistant's voice on user reactions and to enrich the literature in digital marketing. Despite the work on the uses of virtual voice-activated assistants (the levers and obstacles to their adoption in particular), or the work on the voice of chatbots and virtual agents, we found no study that was specifically interested in the voice of the assistant. If the voice gives substance to a machine, and more particularly to a voice assistant devoid of visual representation, we are faced with two possible choices: to provide the assistant with a synthetic voice that resembles a human voice in every way or to opt for a synthetic voice which, although based on human phonemes, allows the user to identify that he is exchanging with a machine, and not with a human. Our qualitative consumer research confirms the literature on the importance of the assistant's voice. On the other hand, our analysis on the voice is opposed to the literature carried out on virtual agents, we highlight that a voice assistant equipped with a more robotic synthetic voice seems better appreciated than an overly realistic human voice. From a managerial point of view, this research aims to suggest lines of thought for designers of voice assistants as well as brands wishing to develop their voice

assistant and improve the relationship with their consumers (Viot and Bressolles, 2012). The voice offers tremendous opportunities for experimentation and invention of new uses, an unprecedented opportunity to create an online experience specific to each site and to stand out on the web (Reeves et al, 2019). The next few years promise a disruption of web browsing inherited from aging HMIs, voice assistants are likely to significantly change both consumer behaviors and marketing strategies (Mari et al., 2020). The voice opens up new perspectives, where the rules of the game have to be reinvented (Kumar, 2018). Given the exploratory nature of our research, our conclusions still have certain limitations. Our study has limited external validity due to the size and composition of our sample. It would be appropriate to undertake an experiment on a voice assistant in order to modulate its voice (human versus robotic) and test the impact of the different voice characteristics on the users.

References

Ammari, T ; Kaye, J. Tsai, J.Y ; Bentley, F. (2019). « Music, Search, and IoT: How People (Really) Use Voice Assistants », ACM Transactions on Computer-Human Interaction 26(3), pp. 1-28. DOI:10.1145/3311956

Chattaraman, V., Kwon, W.-S., Gilbert, J.E., Ross, K. (2019). « Should AI-Based, conversational digital assistants employ social- or task-oriented interaction style? A task-competency and reciprocity perspective for older adults », Computers in Human Behavior 90, 315–330. https://doi.org/10.1016/j.chb.2018.08.048

Chérif, E. et 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, Janvier 2019, vol. 34, no. 1, pp. 29-49.

Cornut, G. (2009). « La voix », Presses Universitaires de France, « Que sais-je ? ». DOI:https://www.cairn.info/la-voix--9782130576747.htm

Devillers, L. ; Masure, A. ; Obin, N. (2020). « Assistants connectés : des ordres dans la voix », interview dans le cadre l'émission de radio Magnétique, Lausanne, Radio Télévision Suisse.

Dubiel, M., Halvey, M., & Azzopardi, L. (2018). « A Survey Investigating Usage of Virtual Personal Assistants », Computer Science ArXiv, abs/1807.04606.

Gouliaéva, O ; Dosquet, E ; Moysan, Y (2020). « La révolution des assistants vocaux : comprendre les enjeux et réussir ses stratégies marketing ». Collection Marketing/Communication, Dunod

Kowalczuk, P. (2018). « Consumer acceptance of smart speakers: a mixed methods approach: An International Journal », Journal of Research in Interactive Marketing 12, 418–431. https://doi.org/10.1108/JRIM-01-2018-0022

Kumar, V. (2018). « Transformative Marketing: The Next 20 Years », Journal of Marketing 82, 1–12. https://doi.org/10.1509/jm.82.41

Lahoual, D. et Fréjus, M. (2018). « De l'utilisabilité à l'appropriabilité des assistants vocaux. Étudier les interactions vocales en situation domestique à partir d'une démarche centrée utilisateurs », Conférence internationale ERGO'IA.

Lee K. M. et Nass C. (2003). « Designing social presence of social actors in human computer interaction », Proceedings of the Conference on Human Factors in Computing Systems - CHI '03, (5), 289.

Lemoine, J.F. et Notebaert, J.F. (2011). « Agent virtuel et confiance des internautes vis-à-vis d'un site web », Décisions Marketing, Janvier 2011, 61, pp. 47-53.

Liao Q., Zhang S., Wang M., Li J., Wang X., Deng X. (2020). « Comparing the User Preferences Towards Emotional Voice Interaction Applied on Different Devices: An Empirical Study », Conference on Human-Computer Interaction. https://doi.org/10.1007/978-3-030-49062-1_14

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. DOI:10.1177/0961000618759414

MacDorman, K.F. (2019). « Masahiro Mori's The Uncanny Valley: Significance and Impact on Aesthetics and Robot Design », Journal of the history of technology. DOI.org/10.4000/ephaistos.5333

Mari A, Mandelli, A. et Algesheimer, R. (2020). « The Evolution of Marketing in the Context of Voice Commerce », 22nd International Conference on Human-Computer Interaction.

Mclean, G. et Osei-Frimpong, K. (2019). « Hey Alexa… Examine the variables influencing the use of Artificial Intelligent In-home Voice Assistants », Computers in Human Behavior (99), pp. 28-37. DOI:10.1016/j.chb.2019.05.009

McLean, G., Osei-Frimpong, K. and 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, pp. 312-328.

Mehrabian A. (1967). « Orientation behaviors and nonverbal attitude communication », Journal of Communication, 16, pp. 324–332.

Nass C. et Moon Y. (2000). « Machines and Mindlessness: Social Responses to Computers », Journal of Social Issues, 56(1), pp. 81–103.

Natale, S. (2020). « To believe in Siri: A critical analysis of AI voice assistants », Communicative Figurations Working Papers, 2020, Vol. 32, pp. 130–146.

Obin, N. et Roebel, A. (2016). « Des voix de synthèse presque humaines », Revue pour la Science, numéro 470, pp. 54-62

Pagani, M ; Racat, M, ; Hofacker, C.F. (2019). « Adding voice to the omnichannel and how that affects brand trust », Journal of Interactive Marketing, volume 48, pp. 89-105. DOI:10.1016/j.intmar.2019.05.002

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 ? », 29ème conférence francophone sur l'Interaction Homme-Machine, Poitiers, France. pp. A-67. hal-01577674

Porcheron, M. ; Fischer, J.E. , Reeves, S. ; Sharples, S. (2018). « Voice Interfaces in Everyday Life », 2018 CHI Conference on Human Factors in Computing Systems, ACM, paper 640, pp. 1-12. DOI :10.1145/3173574.3174214

Qiu L. et Benbasat I. (2009). « Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems », Journal of Management Information Systems, 25(4), pp. 145–182.

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.

Rzepka C., Berger, B. et Hess, T. (2020). « Why Another Customer Channel ? Consumers' Perceived Benefits and Costs of Voice Commerce », 53rd Hawaii International Conference on System Sciences.

Shen, J. ; Pang, R. ; Weiss, R.J ; Schuster, M. ; Jaitly, N. ; Yang, Z. ; Chen, Z. Zhang, Y. ; Wang, Y. ; Skerry-Ryan, R.J ; Saurous, R.A ; Agiomyrgiannakis, Y. ; Wu, Y. (2018). « Natural TTS synthesis by conditionning wavenet on mel spectogram predictions », Google, Inc., University of California, Berkeley.

Stern S.E., Mullennix J.W., Dyson C. et Wilson S.J. (1999). « The persuasiveness of synthetic speech versus human speech », Human Factors, 41, pp. 588–595

Sundar, S. S., Jung, E. H., Waddell, F. T. and 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, pp. 88-97. DOI:10.1016/j.ijhcs.2016.08.006

Velkovska J. et Zouinar M. (2018). « The illusion of natural conversation: interacting with smart assistants in home settings », 2018 CHI Conference on Human Factors in Computing Systems, ACM, DOI:10.1145/3170427.3170619

Viot, C. et Bressolles, G. (2012). « Les agents virtuels intelligents : quels atouts pour la relation client ? », Décisions Marketing, (65), pp. 45-56.

Wang W. et Benbasat I. (2007). « Recommendation Agents for Electronic Commerce: Effects of Explanation Facilities on Trusting Beliefs », Journal of Management Information Systems, 23, pp. 217-246.