Ergonomics In Design, Vol. 77, 2023, 150–158
https://doi.org/10.54941/ahfe1003372

Using Kansei Design to Create a Predictive Model for Voice User Interfaces for Electronic Appliances

Francisco Rebelo[1,2], Maria Mendonça[3], Ricardo Dias[3], João Costa[3], Elisangela Vilar[1,2], and Paulo Noriega[1,2]

[1] CIAUD, Research Centre for Architecture, Urbanism and Design, Lisbon School of Architecture, Universidade de Lisboa, Portugal
[2] ITI/LARSyS, Universidade de Lisboa, Portugal
[3] Lisbon School of Architecture, Universidade de Lisboa, Portugal

ABSTRACT

Nowadays, virtual voice assistants, are present in various devices – from smartphones to smart speakers and, soon, in all electronic appliances. In this context, it is a significant increase in the user experience with them to make them more engaging. Virtual assistants usually have the same voice tone to give users information. Thus, companies are not fully taking advantage of voice properties and the power they hold on to communication, increasing the user experience and the market product success. In this study, we used the Kansei Design method to predict the emotional user reaction impact of different voice user interfaces for electronic appliances. Resorting to some literature review and online post-research, we defined voice characteristics to manipulate – gender, cadence, and inflection. Related to the semantic space (Kansei words) of the user's perceptions with virtual assistant voices – pleasure, proximity, and arousal. Eighty-three participants (67,5% female and 32,5% male) answered an online questionnaire with 12 possible combinations between gender; cadence; and inflection, with the Kansei words: pleasure; proximity; and arousal. Results vary with the semantic space, but they suggest that people felt more comfortable and relaxed hearing a male voice than a female one. Regarding cadence, a typical speech flow was where people felt more intimate with the voice. Though, participants felt more activated while hearing a female voice, speaking at a higher speed and with inflection in her voice. We generated some use cases with these results to understand how they can guide design processes regarding voice-user interfaces.

Keywords: Voice assistance, Smart devices, Kansei design, Predictive models

INTRODUCTION

Nowadays, virtual voice assistants, such as Alexa, Siri, or Bixby, are substantially increasing. They are an artificially intelligent virtual agent that doesn't have a physical representation and is embedded in devices, such as mobile phones and smart speakers. Typically, voice assistants are activated by a keyword, and the digital assistant interprets and return information or performs a specific function without requiring a hand interaction with the device.

Smart devices that use virtual voices are designed to have a natural conversation with an authentic human voice, avoiding the robotic voices that can lead to problems with user perceptions and trust. Nass and Moon (2000) showed that people mindlessly apply the same social norms used for interpersonal communications to interact with devices, such as applying social categorizations like "male," "female," or "trustworthy." In this context, the virtual voice is needed to speak information that contains cues, including vocal pitch, that may be unconsciously utilized for social perceptions, like what is found in human communication (Oleszkiewicz et al., 2017).

In this study, we consider that persons interact with smart systems in the same way as with real people. Unconsciously, users apply the same social rules used in interpersonal communication and view computers as a peer in social interaction (Nass et al., 1996; Reeves & Nass, 1996).

Previous studies have highlighted factors including: social presence, affecting the outcomes between virtual assistance and humans in services contexts (Etemad-Sajadi, 2016); trust in the service provider (Hess et al., 2009); satisfaction with the service encounter (Verhagen et al., 2014) and communication, recommendation quality, and improved customer motivation and engagement (Baylor, 2011). In this study, we focus on the influence of the voice characteristics present in smart speakers on people's engagement and presence.

The main objective is the creation of a predictive model to help interaction designers to propose the voice characteristics for smart electronic appliances, in the function of user engagement and presence. To develop this predictive model, we use Kansei Design, as a support to get and systematize data acquired in a human-centered process (Nagamachi, 1995).

METHODS

There are five main steps in applying Kansei Engineering in a product development process:

i. Defining the domain of study;
ii. Describing the domain according to: a. Semantic Scale (involves an extensive research process that includes case study analysis and literature review to find out a large vocabulary of keywords on the subject, that are then grouped in an affinity map); b. Product Specification (involves deconstructing the product to its essential characteristics and defining the possible variations of these specifications);
iii. Synthesizing a set of product properties and Kansei Words;
iv. Validating the semantic space and space of application;
v. Building a model, i.e., an inference motor that will analyze all the permutations of the product specifications and compare them to the answers given by users.

The domain description in this project is related with the selection of the voice assistance characteristics, for smart electronic appliances, in the function of the user engagement and presence.

The development of the semantic space and the space properties are supported by literature review and online posts. Surprisingly, we did not find a considerable amount of information, especially in literature. Most of the online posts found were about gender bias and why computer voices are mainly female. Literature also studied differences on how female voices are perceived, comparing it to male voices. Another part of online posts talked how smart assistant voices can sound robotic and how should we make them feel more human. One last part of our research found some contributions in speech properties and contribution to vocal emotion and virtual assistant's voice.

With the semantic space and product properties defined, we then created a questionnaire to assess how people feel according to each voice. For the purpose, we opted for Google Forms. The first section of the questionnaire introduced people to the main research objectives and how to answer to the questionnaire. To each voice, a short video was created: recurring to online text-to-speech, we generated the 12 possible combinations saying a short, informative phrase: Today the weather will be rainy with a chance of thunder at 12pm. All the videos have the same background: a smart speaker lighting up, indicating some triggering action just happened. The questionnaire had both term of consent and questions translated to Portuguese.

The Kansei questionnaire applied has a sample of 83 participants, 67.5% are female and 32.5% male, with a mean age of 32.6 years and a standard deviation of 13.9. Regarding the professional status of the participants, there is a higher prevalence (49.4%) in the employed response. 54 participants (65,9%) declared not to use virtual assistants, while 28 of them do (34,1%).

RESULTS AND DISCUSSION

Results From Semantic Space, Kansei Words

We collected a total of 29 words (27 unique words) related to user experience and how people feel with different types of voices: Irritation; Anxiety; Sadness; Happiness; Pleasure; Anger; Fear; Despair; Elation; Natural; Kind; Natural; Credible; Rational; Persuasive; Competent; Likeable; Soothe; Calm; Friendly; Flirty; Robotic; Formal; Casual; Friendly; Efficient; Cold; Chatty; Witty.

To create different semantic scales, we proceeded, resorting to a manual expert method (affinity diagram), to link the different words into categories for each word. Table 1 shows the results generated from this method, 3 Kansei words (Pleasure, Proximity, Arousal) and its semantic scale.

Each scale, in which we define three different factors to study, is related to how a person would feel hearing each voice:

- Pleasure – we wanted to measure if people liked the voice they were hearing - if the voice makes them uncomfortable or not. This word measures the user experience level.
- Arousal – our goal is to measure whether people might feel calmer or more excited while hearing the assistant speaking. Like pleasure, this word measures the user experience level.

Table 1. Kansei words, affinity analysis result.

Pleasure		Arousal		Proximity	
Uncomfortable	**Cosy**	**Calm**	**Excited**	**Impersonal**	**Pessoal**
Happiness	Likeable	Despair	Anxiety	Rational	Credible
Kind	Flirty	Despair	Sadness	Formal	Cold
Chatty	Casual	Fear	Anger	Soothe	Efficient
Friendly	Elation			Robotic	

- Proximity – in our study, measuring proximity relates to how people feel the voice regarding an intimate level – if they feel in a close relationship with a voice or if they feel distant.

Results From the Semantic Space

With this process, we delineated three aspects to change in the voices:

- Gender – there are a few vocal differences related to gender, but, while setting this characteristic for the study, our motivation was to test and study pitch. Men tend to speak with a lower pitch than Women do. While literature shows that higher pitch voices tend to make people feel nervous or emotionally unstable, most of the commercial smart-speakers, at this moment, usually have a distinctly female voice rather than a male voice (being the lower pitch related to higher levels of credibility, competence, and attractiveness).
- Cadence – this property is related to differences in our speech flow. We can opt for a higher cadence – faster rhythm – while talking or a lower cadence – slower rhythm. The pace at our words take is important for the message to be clear. Literature suggests that fast speech is linked to persuasive and credible speakers.
- Inflection – voice inflection is an important way to express feelings. This property corresponds to how people intonate different parts of a phrase to emphasize the message they want to pass, changing how others can perceive it. The speech is perceived as monotonous and flat if there is a lack of intonation. After research, we delimited our three properties and their variations, having a total of 12 different combinations (Table 2) to work with in our questionnaire.

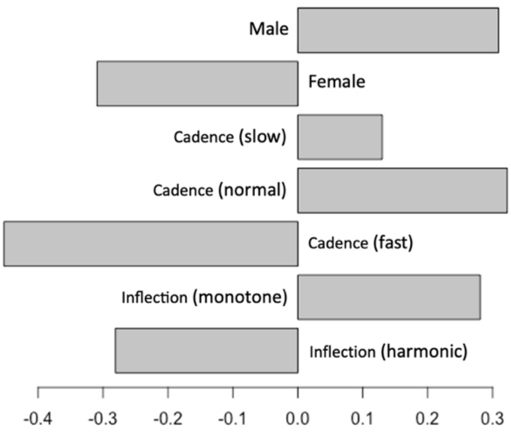
To process the data we used the Quantification Theory Type I (QT1), a multiple regression analysis that works with dummy variables and with the average of all individual responses.

Graphs 1 to 3 show the results of the QT1 test, related to the link between voice properties (Gender, Cadence, Inflection) for the Kansei evoked words

Table 2. Voice properties.

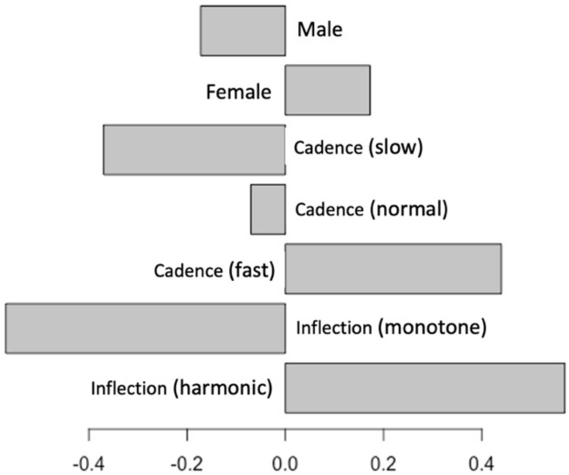
Gender	Cadence	Inflection
Male	Slow	Monotonous
Female	Normal	Harmonious
	Fast	

(Pleasure, Activation, Presence). In the graphs, a positive score implies positive emotional vocabulary, and a negative score implies negative emotional vocabulary.



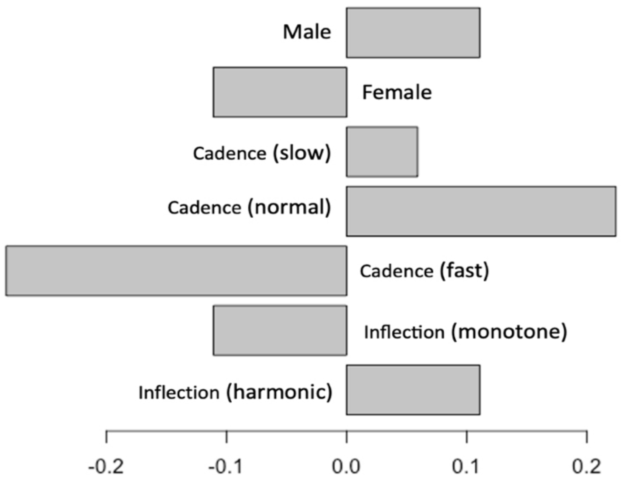
Graph 1. Preferences between pleasure and voice properties.

The results in Graph 1 show that the variables influencing the perception of pleasure more positively are the male voice, with a normal cadence and a monotone inflection. On the contrary, the variables negatively influenced the perception of pleasure was the female voice, with a fast cadence and a harmonic voice inflection.



Graph 2. Preferences between activation and voice properties.

Results for Graph 2 show that the variables influencing the perception of activation more positively are the female voice, with a fast cadence and a harmonic inflection. On the contrary, the variables negatively influenced the perception of pleasure was the male voice, with a slow cadence and a monotone voice inflection.



Graph 3. Preferences between proximity and voice properties.

For the keyword proximity, the variables influencing the perception of pleasure more positively are the male voice, with a normal cadence and a harmonic inflection. On the contrary, the variables negatively influenced the perception of proximity was the female voice, with a fast cadence and a monotone voice inflection. We particularly draw attention to the perception of pleasure and proximity with the male voice and the perception of activation for the female voice. These previous results contradict what currently occurs in the market, which uses the default female voice, called by a feminine name, Alexa from Amazon, Siri from Apple, and Cortana from Microsoft. In the particular case of Google, which did not use a human name and selected a female voice.

Regarding this option, LaFrance (2016), argues that academics and developers made this decision due to cultural expectations. Women traditionally assume administrative roles in families and are considered friendly. This could decrease fear about the technology of virtual assistants' power, thus avoiding rejection.

The fact that we found this unexpected result led us to ask whether this result would be influenced by gender. , show the shows gender preferences

Table 3. Gender preferences regarding Kansei words for voice features.

	Gender effect in male or female voice		**Gender effect in voice cadence**		**Gender effect in inflection**	
	Male Sample	Female Sample	Male Sample	Female Sample	Male Sample	Female Sample
	Preferences					
Pleasure	Male	Male	Normal	Normal	Monotone	Monotone
Activation	Female	Female	Fast	Fast	Hamonic	Hamonic
Proximity	Male	Male	Normal	Normal	Hamonic	Harmonic

Table 4. A predictive model for the voices properties (Kansei model).

	Voice Properties						
	Gender		Cadence			Inflection	
Kansei Words	Male	Female	Slow	Normal	Fast	Monotone	Harmonic
Pleasure	+ +	- -	+	+ +	- - -	+ +	-
Proximity	+ +	-	+	+ + +	--	-	+ +
Activated	- -	+	- - -	-	++	- - - -	+ + +

for the Kansei words pleasure, activation, and proximity for the manipulated voice characteristics (voice gender, cadence, and inflection).

Regardless the gender, we see the same trend, the male voice positively influences pleasure and proximity, and the female voice influences activation. The same trend can be seen concerning voice cadence and voice inflection, where there are no differences between genders.

Considering those results, we propose a personalized solution for the predictive model (Table 4) for voices characteristics.

So, in order for the user to feel cozy (pleasure), our voice would be male, speaking at a normal rate and with a monotone inflection. To feel close to the voice, the user will prefer the same combination: a male voice, speaking at a normal rate and with a monotone voice. Though, the user will feel more excited (activation) if he hears a female voice speaking at a fast rate and with a harmonious inflection.

To materialize the results from the Kansei model, four fictional situations were created, and properties of a VUI were applied in order to create specific feelings in users, accordingly.

Use Case 1 - Waking Up

Brief Description: Waking up is a delicate part of the day-to-day life of any adult. In this scenario, our persona, Mário, has to attend a meeting earlier than normal in his job. As such, he asks his personal assistant to wake him up at 6 am in the following morning. As asked, at 6 am, the Virtual Assistant wakes up Mário by telling him: "Good morning Mário. It's six am on this beautiful morning. Your meeting is in two hours in the Lisbon office."
Intended Feeling: Cozy.
Voice Properties: Male, Normal, Monotone.

Use Case 2 - Fire Detected in the Building

Brief Description: Because he is a very busy man, Mário is usually a multitasker. This means that sometimes his attention is divided between working at the beginning of the evening and cooking dinner. One night, Mário left his oven on while he went to his laptop to answer emails. Because a cloth was too close to the oven, it caught fire, burning part of the kitchen and turning on the smart smoke alarm. The VUI in Mários' smartwatch says: "Mário, there is a fire in the kitchen, so everyone must leave the house immediately."
Intended Feelings: Excited.
Voice Properties: Female, Fast, Harmonic.

Use Case 3 - Telling a Child a Story to Sleep

Brief Description: Mário's 17-year-old nephew spent a night at Mário's house. Because he was not used to sleeping there, he had difficulties falling asleep. Because he is very intrigued by technology and smart speakers, Mário asks the voice assistant to tell his nephew a story to sleep. The VUI responds: "Ok, this one is called Little Red Hood. Once upon a time…"
Intended Feelings: Calm.
Voice Properties: Slow, Monotone, Male.

Use Case 4 - Detecting Trespassers in the House

Brief Description: One night, Mário was out for dinner with his family. The house was empty, and, as such, the smart alarm with a movement detector was on. Around midnight, the alarm detects people in the house. After attempting to check the identity of the people without success, the VUI alerts the intruders: "The identity of the people in the house were not verified. Competent authorities were informed about the trespassing and are on the way."
Intended Feelings: Uncomfortable.
Voice Properties: Fast, Harmonic, Female.

CONCLUSION

This study aimed to create a predictive model to help interaction designers propose the voice characteristics for smart electronic appliances in user engagement and presence. For this purpose, we used the Kansei method to develop a semantic scale with the words representing the reactions that a sample of participants reported concerning the combination of assistance voice properties (gender, consonance, and inflection). The results showed a preference for the male voice for the Kansei words pleasure and proximity and a preference for the female voice for the Kansei word activation.

The preference for the male voice does not meet the trends seen in the market, which adopted the female voice for virtual assistants. To see if there is any gender effect on this preference, we apply Kansei to both the male and female samples. The results showed the same trend as those found in a mixed sample.

From the creation of a predictive model, which relates the properties of the voice with the reactions (pleasure, activation and proximity). For this model, four interaction scenarios were created to demonstrate its use.

Some limitations can be highlighted, and we make recommendations for future investigation. First, our sample could be more homogeneous and representative of different populations. Since we have around 67.5% of female respondents and the sample comes from Portugal. Future studies should have a greater representation of other countries, with other cultures, and with a more robust and homogeneous sample.

In the end, this study provided an initial contribution for literature reviews around the topic of computer voices and smart speakers, a topic we believe could be more studied and guided for future design work since there are a lot of studies of how voices and the way it sounds affects political or commercial communication.

ACKNOWLEDGMENT

The authors would like to acknowledge the Portuguese National funds finance this work through FCT - Fundação para a Ciência e a Tecnologia, I.P., under the Strategic Project with the references UIDB/04008/2020 and UIDP/04008/2020 and ITI-LARSyS FCT Pluriannual fundings 2020- 2023 (UIDB/50009/2020).

REFERENCES

Baylor, A. L. (2011). The design of motivational agents and avatars. Educational Technology Research & Development, 59(2), pp. 291–300.

Etemad-Sajadi, R. (2016). The impact of online real-time interactivity on patronage intention: The use of avatars. Computers in Human Behavior, 61, 227–232.

Hess, T. J., Fuller, M., & Campbell, D. E. (2009). Designing interfaces with social.

presence: Using vividness and extraversion to create social recommendation agents. Journal of the Association for Information Systems, 10(12), 1.

LaFrance, A., 2016, March 30. Why do so many digital assistants have feminine names? The Atlantic. https://www.theatlantic.com/technology/archive/2016/03/wh y-do-so-many-digital-assistants-have-feminine-names/475884/.

Nagamachi M. (1995). Kansei Engineering: a new ergonomic consumer-oriented technology for product development. Int. J. Ind. Ergon., 15, 3–11.

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? International Journal of Human-Computer Studies, 45(6), 669–678.

Nass, C., Moon, Y., 2000. Machines and mindlessness: social responses to computers. J. Soc. Issues 56 (1), 81–103. https://doi.org/10.1111/0022-4537.00153.

Oleszkiewicz, A., Pisanski, K., Lachowicz-Tabaczek, K., Sorokowska, A., 2017. Voice-based assessments of trustworthiness, competence, and warmth in blind and sighted adults. Psychon. Bull. Rev. 24 (3), 856–862. https://doi.org/10.3758/s13423-016- 1146-y.

Reeves, B., & Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people. Cambridge, United Kingdom: Cambridge university press.

Verhagen, T., Van Nes, J., Feldberg, F., & Van Dolen, W. (2014). Virtual customer service agents: Using social presence and personalization to shape online service encounters. Journal of Computer-Mediated Communication, 19(3), 529–545.