

---

# Essential factors for Emergent Users in Smart Speakers

**Shashank Ahire**

Human-Computer Interaction  
Leibniz University Hannover  
Hannover, Germany  
shashank.ahire@hci.uni-  
hannover.de

**Michael Rohs**

Human-Computer Interaction  
Leibniz University Hannover  
Hannover, Germany  
michael.rohs@hci.uni-  
hannover.de

**Abstract**

Smart speakers are increasingly adopted by diverse user groups across globe. One such user group rapidly adopting the latest technology is emergent user group. Emergent users generally have minimal formal education, less technological experience, and are resource constrained. In our previous research, we designed a smart speaker for emergent users and combined human and AI response to minimize error rate and deliver efficient and elaborate response to them. Based on our observations and findings in this study, we found essential factors for emergent users in smart speaker design and interaction. We recommend the following ten factors; multilingualism, accents and dialects, data consumption, energy consumption, useful and functional use cases, design and space, mental model, cognitive load during an interaction, error tolerance, and accessibility to different demographics and gender.

**Author Keywords**

emergent users; smart speakers; design; factors; conversational user interface

**Introduction**

Smart devices are mainly designed and developed in the “first world”<sup>1</sup>. These devices are build focusing mainly on population based in “first world”. However, these devices

---

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Copyright held by the owner/author(s).  
CHI'23, April 23–28, 2023, Hamburg, Germany

---

<sup>1</sup>[https://en.wikipedia.org/wiki/First\\_World](https://en.wikipedia.org/wiki/First_World)

are sold and adopted across the globe. Similarly smart speakers are designed and developed based on the requirements of users located in the west. Furthermore, they are also trained on data generally originating from native English language speakers. The phrase “*Designed in California*” has been extensively used on smart devices that are meant to be consumed by different types of consumers. The centralization of design, development, and training data has created problems for different types of user groups, user groups that are diverse with respect to culture, language, lifestyle, and adoption of technologies. Although, it is claimed that smart speakers enable user to perform their tasks effortlessly, but several studies have found that smart speakers are struggling with recognizing user requests, interpreting it, and generating a satisfactory response.

Smart speakers are increasingly penetrating in different geographies, user groups, ages, and languages. Unlike other interfaces, Conversational User Interface (CUI) do not require formal learning. The advantages of CUIs are well known: They enable hands-free and eye-free intuitive interaction and provide swift input. Because CUIs do not require literacy, they have been used as an input method in studies with low-literate user populations.

Emergent users are less educated, have lower English language skills, are often economically disadvantaged, geographically dispersed, and are culturally heterogeneous [3]. Emergent users are located not only in developing economies but also in the world’s developed economies. Emergent users are rapidly adopting mainstream applications like WhatsApp, Facebook, Google Pay, Amazon, and Ola (taxi booking service). Even though these applications are English-dominant, emergent users are adopting these applications because of network effects, hassle-free experience, and most importantly – time, effort, and money-

saving opportunities. Although these applications are not designed and implemented for them, emergent users learn to handle those applications with the help of family, friends, and acquaintances. Furthermore, emergent users are increasingly adopting the latest technologies, such as smartphones, smartwatches, and smart speakers. However, no smart devices have been designed and built considering their specific requirements.

In this position paper, we present essential factors to be considered in smart speaker design for emergent users. Our recommendations are based on findings and observations we made in our previous study [1], in which we designed a smart speaker for emergent users. We initiated our study by identifying emergent user criteria for smart speakers. We evaluated two variants with emergent user family for the period of 14 days. One variant was based on AI-enabled responses and the other based on human responses. The participants of our study judged a smart speaker based on three factors: the language of the request, the length and complexity of the request, and the urgency of response.

Subsequently, we conducted two workshops to determine the design requirements of the emergent users. Based on these findings we built a ‘Human plus AI’ smart speaker. From the results of the final deployment, we conclude that integrating the two concepts proved beneficial to emergent users. The ‘Human plus AI’ smart speaker provided instant, elaborate, and precise responses to emergent users.

## Essential factors

Following, we discuss ten factors that we recommend following while designing a smart speaker.

#### *Multilingualism*

On several instances multilingual users tend to mix two languages or borrow some words from another language while speaking. This habit of borrowing also gets reflected while interacting with smart speakers. Since the smart speaker has no predefined protocol for conversational interaction, it is crucial to provide a buffer to inculcate the mixing of languages. As smart speakers are predominantly trained to recognize English, their recognition accuracy is often lower for other languages.

#### *Accent and dialects*

As emergent users may have different dialects and accents, it is essential that smart speakers are capable of understanding their language and dialect. To achieve this, it is crucial to train the voice recognition model of the smart speaker on the data collected from those users. Training voice recognition model on the data from emergent user will reduce the error rates drastically. Lower error rates of voice recognition model will increase the acceptability of smart speakers among emergent users.

#### *Energy consumption*

Energy consumption is one of the significant concerns in emergent users households. When buying electronic appliances, power efficiency and energy ratings are major factors [4]. The energy saving of the object indicates that the appliance is energy efficient. Energy efficiency is especially crucial if an electronic gadget functions  $24 \times 7$ .

#### *Data consumption*

Currently, smart speakers are entirely dependent on the internet for voice recognition. Therefore, a reliable internet connection is a prerequisite for a smart speaker. The process of recognizing a request and delivering a response should consume minimal data for processing the request. Data usage is an essential factor for smart speakers. Higher

data consumption could discourage emergent users from using the smart speaker. Increased data usage could even lead to the abandonment of a smart speaker. Smart speakers are capable of executing tasks that could consume large amount of data such as streaming music and video, playing games, or making a video call. Therefore, smart speakers should have a provision to set a limit, or monitor its usage.

#### *Useful and practical use cases*

Currently, smart speakers can perform tasks like, entertainment, Internet search, reminders and scheduling, and smart home control. Although all of these tasks are important but are not crucial, and there are alternative ways to accomplish them. To ensure that smart speakers are helpful to emergent users, it is essential to assist them in activities crucial to their day-to-day work. For instance, as emergent users have pursued limited education in their regional language, they struggle to help their children with their education. Smart speakers should be capable of assisting emergent users in their children's education.

#### *Design and space*

The design of the product is an important consideration in the buying process. The design builds a connection and helps us to connect to the product. Replicating and selling the same design across different countries has been a strategy of companies like Google, Amazon, and Apple. Often smart speakers are designed to appeal to households located in western countries. Hence, design decisions are largely influenced by western perspectives. On the other hand, households of emergent users are completely different. They are compact and unstructured. Physical space that is consumed by the component is a crucial consideration when selecting an object, as the required amount of space will determine its location in the house. Hence, if the

smart speaker consumes enormous space, this could lead to placing it in an inconvenient or inaccessible location. The suboptimal placement could entail numerous voice recognition problems.

#### *Mental model*

The mental model is an essential aspect during a smart speaker interaction. An easy-to-build mental model of a smart speaker interaction will foster a seamless experience. The current mental model of smart speaker interaction is designed for mainstream English speakers. It is difficult for an emergent user to frame their mental model based on current interaction design. Furthermore, the smart speaker has no visual cues to help users, like guiding them on initiating an interaction, telling them how to use it, or explaining its capabilities. Hence, it is difficult for users to initiate an interaction based on intuition or previous experience.

Suppose the mental model turns out to be a hindrance for interaction. In that case, it is difficult for emergent users to initiate a smart speaker interaction and consistently interact with it for the long term. Smart speaker companies rarely provide tutorials that could guide users about smart speaker functionalities.

#### *Cognitive load during interaction*

It has been found that during a voice interaction, the cognitive load of non-native speakers vs. native speakers is higher [5]. As cognitive load for translating to English language, structuring the request in the correct English sentence formation requires high cognitive energy. It is effortful for non-native language speakers to interact with smart speaker.

Moreover, due to the hand-free and eyes-free features of the smart speaker, users tend to dual-task while interacting with them. The cognitive load while performing differ-

ent tasks during interaction is significantly high [2]. While designing a smart speaker for emergent users, it is essential to consider cognitive load during regular and dual-task interaction. The minimal cognitive load during a smart speaker interaction will increase the number of interactions with smart speakers and motivate emergent users to interact with them consistently.

#### *Error tolerance*

Error tolerance in user interfaces is an important aspect. Error tolerance helps users to build confidence during the interaction and encourages them to interact with the system. Smart speakers have a relatively low error tolerance. If a user commits a mistake during an interaction, it typically delivers an error response. To complete a task, the user has to try again, often multiple times. The low error tolerance leads to a poor user experience.

For instance, smart speakers have a low error tolerance for grammatically incorrect sentences. Smart speakers should be capable of providing a useful response even for sentences structured incorrectly. For a grammatically incorrect request, the smart speaker's language model should try to restructure the request automatically. Based on the request, it should deliver an answer rather than delivering only an error response. A higher tolerance for grammatical errors will encourage emergent users to try an interaction when they have low confidence or when they are unsure of what they seek. A higher tolerance for errors could help users gradually build confidence for interactions. This could help them foster long-term engagement and skill development.

#### *Accessible to gender and different demographics in a family*

In our study, we observed that the adoption of smart speaker in emergent user houses was hierarchical. For the first few days, the male adult member of the family learns to operate the smart speaker. After learning how to operate the

smart speaker, he teaches it to the teenage family member, who are generally technologically savvy. Later, it is taught to female adult members of the family. Typically female adult members have less technological exposure. Due to fear of committing a mistake or damaging the device, family members with low level of technical knowledge refrain from using the device.

In a hierarchical structure of technological adoption, the smart speaker must be accessible to the members at the lower levels of technological adoption. The accessibility will remove the fear of committing a mistake and build confidence to play and explore smart speakers on their own. Therefore, a smart speaker should be accessible to all the members of the family irrespective of their gender, age, or technological experience.

## Conclusion

As emergent users increasingly adopt the latest technologies, it is essential to consider their requirements in designing and developing those technologies. In this paper, we presented ten important factors for emergent users in the smart speaker. The essential factors are, multilingualism, accent and dialects, energy consumption, data consumption, useful and functional use cases, design and space, mental model, cognitive load during interaction, error tolerance and accessible gender, and different demographics. We appeal to all the researchers and practitioners in the domain of smart speaker and CUI to inculcate these factors in their design and development process.

## REFERENCES

- [1] Shashank Ahire. 2023. Designing a Smart Speaker for Emergent Users: Human Plus AI Response. In *Proceedings of the 13th Indian Conference on Human-Computer Interaction (IndiaHCI '22)*. Association for Computing Machinery, New York, NY, USA, 67–72. DOI : <http://dx.doi.org/10.1145/3570211.3570217>
- [2] Shashank Ahire, Aaron Priegnitz, Oguz Önbas, Michael Rohs, and Wolfgang Nejdl. 2021. How Compatible is Alexa with Dual Tasking? — Towards Intelligent Personal Assistants for Dual-Task Situations. In *Proceedings of the 9th International Conference on Human-Agent Interaction (HAI '21)*. Association for Computing Machinery, New York, NY, USA, 103–111. DOI : <http://dx.doi.org/10.1145/3472307.3484165>
- [3] Devanuj Balkrishan, Anirudha Joshi, Chandni Rajendran, Nazreen Nizam, Chinmay Parab, and Sujit Devkar. 2016. Making and Breaking the User-Usage Model: WhatsApp Adoption Amongst Emergent Users in India. In *Proceedings of the 8th Indian Conference on Human-Computer Interaction (IndiaHCI '16)*. Association for Computing Machinery, New York, NY, USA, 52–63. DOI : <http://dx.doi.org/10.1145/3014362.3014367>
- [4] Crompton. Accessed 2023. What is BEE Rating? <https://www.crompton.co.in/blogs/home-appliances/what-is-bee-rating/>. (Accessed 2023).
- [5] Yunhan Wu, Justin Edwards, Orla Cooney, Anna Bleakley, Philip R. Doyle, Leigh Clark, Daniel Rough, and Benjamin R. Cowan. 2020. Mental Workload and Language Production in Non-Native Speaker IPA Interaction. In *Proceedings of the 2nd Conference on Conversational User Interfaces (CUI '20)*. Association for Computing Machinery, New York, NY, USA, Article 3, 8 pages. DOI : <http://dx.doi.org/10.1145/3405755.3406118>