

User Experience Evaluation of Voice Interfaces: A Preliminary Study of Games for Seniors and the Elderly [†]

Andrés Solano ^{*}, Gustavo López, Luis Guerrero and Luis Quesada

Faculty of Engineering, Computer Science School, University of Costa Rica, 11501, San José, Costa Rica; gustavo.lopez_h@ucr.ac.cr (G.L.) luis.guerrero@ecci.ucr.ac.cr (L.G.); luis.quesada@ecci.ucr.ac.cr (L.Q.)

^{*} Correspondence: andres.solanoportuguez@ucr.ac.cr

[†] Presented at the 13th International Conference on Ubiquitous Computing and Ambient Intelligence UCAmI 2019, Toledo, Spain, 2–5 December 2019.

Published: 21 November 2019

Abstract: Maintaining high cognitive activity is vital for the mental health of seniors. Taking part in leisure activities is an exciting and gratifying way to accomplish this, with video games having several advantages for this use. The present study is a work in progress to develop different digital games with natural voice interfaces that can be easily and pleasantly used by older adults to stimulate their cognitive needs. A Wizard of Oz with two different games was developed to evaluate the user experience perceived when using such applications. Positive results enhance the notion that using digital games powered by natural voice interfaces may appeal to the elderly, while stimulating their cognitive processes, thus fortifying their mental health.

Keywords: voice interfaces; smart-speaker; user-experience; Wizard of Oz; elderly; mental health

1. Introduction

Mental health is a critical factor for maintaining a high quality of life in the aging population [1,2]. However, it is also a health issue, with 50% of elderly people in hospitals presenting some kind of cognitive disease [3]. Clinical studies show that cognitive diseases, delirium, and dementia, mainly caused by Alzheimer’s disease [4], can all have a common period of cognitive decline [2] before being diagnosed. Therefore, it becomes vital to identify what factors can protect a person from this decline to effectively prevent it, therefore promoting their mental health.

However, signaling out these factors may not be enough to make them effective. For example, people with higher levels of education have been identified as being more resistant to the effects of dementia [2]. Having access to this education is a factor that, while effective, cannot be made universal, since worldwide there are no countries where more than 35% of its population finished tertiary education in 2010 [5]. With this in mind, we try to find more applicable protective or preventive factors, which increase the cognitive reserve, thus reducing the cognitive decline in elderly people and lowering the risk of mental health diseases.

Observational studies have shown that exposing elderly people to leisure activities has a positive effect on the protection against cognitive decline [1–4,6,7]. Verghese et al. in their 21-year study demonstrated a clear association between a decrease in the risk of dementia and participating in regular leisure activities [2]. This supports the notion that mental health is enhanced through exposure to leisure cognitive activities.

These activities include a wide range of tasks that aim to stimulate thinking and memory cognitive processes [1–4,6]. Among them, the discussing of past and present events of interest, word games, video games, puzzles, and trivia can be found [6].

Of particular interest in this topic is the use of digital games. Growing evidence supports the idea that video games help seniors meet their needs for fun and mental stimulation [7–9]. Their use in the boosting of cognitive functions in aging populations has advantages given their inexpensive prices, easy access, and the gratification they provide [7]. Moreover, the use of video games as cognitive enrichment tools was identified as an active area of research by Plaza et al. in their literature review [7].

In light of the above, our objective is to develop smart speaker digital games with an optimal user experience that appeal to seniors and aim to stimulate their users’ cognitive processes through leisure activities that can be performed on their own, therefore enhancing their mental health.

To determine if the idea would be well received by the general population, a sketch, seen in Figure 1, was made to evaluate the proposal. The storyboard was evaluated by 10 people, obtaining positive responses and feedback from all participants.

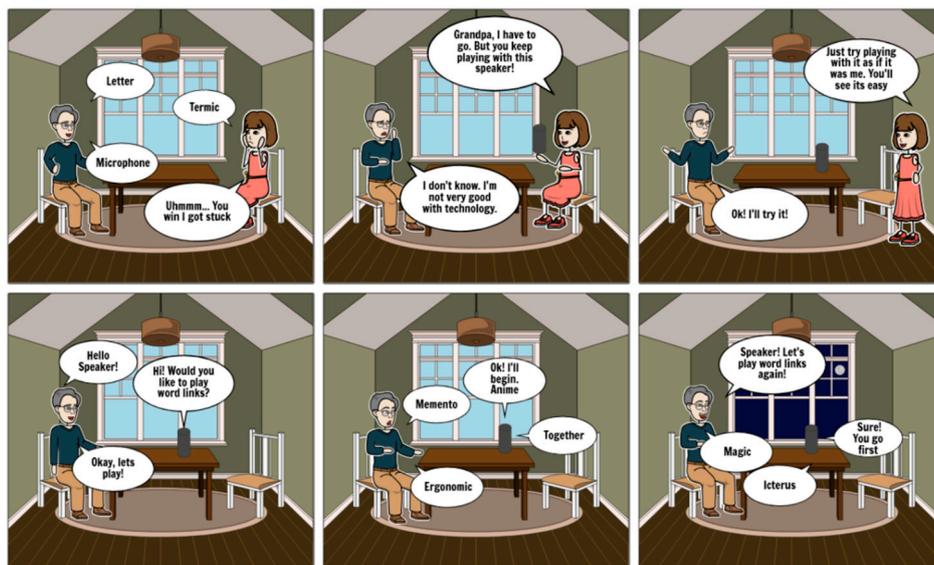


Figure 1. Initial sketch to demonstrate the proposed idea and asses its perception.

Taking this response into account, work began to develop different digital games with natural voice interfaces that can be easily and pleasantly used by the elderly. This paper discusses the first approach, through a Wizard of Oz prototype with two cognitive digital games developed, to evaluate the user experience of the general population against this type of games. Evaluation with the senior population has not been done yet because we believe that work must be done to improve the user experience of these interfaces before testing them with our final users. Even so, positive feedback was received from the present evaluation, which enhances the notion that these interfaces will provide a positive impact on the mental health of the elderly.

The remainder of the paper is structured as follows: Section 2 identifies related works to recognize suggestions to implement smart speaker games for the elderly. Section 3 describes the methodology used of the study, while Sections 4 and 5 present the results obtained and their discussion respectively. In Section 6, conclusions from these results are drawn.

2. Related Work

2.1. Designing Digital Games for Elderly People

Digital games offer various benefits to the elderly, both from an entertainment and a therapeutic point of view [8]. However, to accomplish an impact and promote its usage, games must offer a gratifying user experience to seniors. Various authors have identified a series of attributes that, when incorporated into a video game, cause a better appeal to the elderly. The following recommendations are the main findings from these studies [7–9]:

1. Minimizing game complexity. Implementing games with minimal, uncomplicated rules makes them more comfortable for elderly people. Simple games are the most enjoyable for the population.
2. Developing popular or previously known games enhances the user experience. Adding technology and new simple rules can make the conventional game more enjoyable and rewarding while removing the complexity of learning a whole new activity.
3. The implemented game must have a substantial perceived benefit for the player. If the senior can identify cognitive stimulation from playing, they are more likely to invest their time and energy into playing the game.
4. Games with the capacity to build an audience are desirable. Involving as many players as possible in the gameplay has the additional benefit of meeting the social and bonding needs of the elderly.

2.2. Creating Positive Experiences with Voice Interfaces

Since spoken language is the most natural and efficient form of communication, voice interfaces have the potential to offer an optimal user experience [10–12]. However, to date, there are large limitations in the available technology to implement these systems [13]. Even so, different authors have studied and described design principles and heuristics to follow when conceptualizing natural voice interfaces [11–13].

Some of the main designed principles were presented as solutions against the limitations of voice interface services to create natural and usable interfaces [12]. This includes taking into account the context of the conversation through using previous responses to infer what the user is trying to express. Being as informative as needed, presenting clear and unambiguous dialog, while maintaining it short [12].

On the other hand, Shum [13] defines 10 heuristics that can be used to evaluate and design voice interfaces. The most relevant ones for the present study are to maintain a controlled number of simple options for the user to handle per interaction, control the amount of speech generated by the interface and offer alternatives on how to receive the user's input. Moreover, providing cautiously designed feedback and using yes and no questions, which are usually robust, is recommended [13].

Finally, to evaluate a voice interface, a Wizard of Oz prototype is recommended by Spiliotopoulos, et al. in their 2009 study [11]. Using a Wizard of Oz allows the evaluator to focus on the functionality and user experience of the system's interfaces, without worrying about the technical limitations and the complexities of an internal implementation needed for a smart voice system or application [11].

3. Materials and Methods

To evaluate the user experience of using leisure games for smart speakers, two different products were evaluated during the first semester of 2019: A Wizard of Oz application built to play against the participants of the evaluation, with two different prototype games. To compare the experience that current smart speaker technology offers with what our prototype offers, an implemented version of one of the prototyped games was also evaluated.

3.1. Materials

The constructed prototype consisted of a local computer connected to a Bluetooth speaker, giving the idea to the user that he or she was interacting with a fully working smart speaker. The voice engine used to answer and interact with the participant through the speaker was the built-in text-to-speech voice from the macOS operating system.

The computer had two separate game applications installed, which gave the wizard conducting the evaluation of all the possible interactions with the user, and speaking through the Bluetooth connected speaker when prompted. Both games prototyped were popular cognitive games that focus on stimulating memory and thinking, with simple rules to follow.

The first of the games is a Spanish version of the game Word Chains, where participants are required to form sequences by saying a word that starts with the last syllable of the word the other player said. The user would start by saying any word they wished. The wizard would type the word into the prototype, which would determine if the word was valid (it followed the sequence and had not yet been said) and speak out loud a correct word that followed the sequence. The user will then answer out loud with another word. This interaction repeated until the user or the wizard could not come up with a word that followed the sequence. Quick interactions were pre-added to the application to inform the player of instructions, winning, and losing the game.

The second application prototyped was a version of the popular Akinator game. To play, the wizard would ask the user to think of a popular character or person. Once he had one in mind, Akinator would ask a series of questions, that the user could answer with the options yes, no or I do not know, to guess the character. If after 30 questions it was unable to guess, the user would win.

The speaker would greet the participant when the wizard prompted it to and would then guide the user to play both games. Starting with Word Chains, it would explain the instructions to the user, and once it confirmed the user understood, it would begin playing with the interaction flow detailed above. Once either the user or the wizard lost, it would ask the user if he or she wished to play another round or proceed to the next game. When the user asked for the next game, the same sequence of actions occurred with Akinator.

Finally, the live application chosen was a version of the Akinator game implemented for Google Assistant, running on Google Home. Since a live version of Word Chain was not found on an existing smart speaking platform, this evaluation was only done with Akinator.

To evaluate the user experience of the participants for both applications, the User Experience Questionnaire (UEQ), a tool for measuring the user experience of interactive products [14], was used.

3.2. Participants

A total of 25 participants, varying in gender and ages ranging from 20 to 58, took part in the evaluation. This was done to obtain initial impressions on the user experience from a group of people with whom it is easier to interact. The next step would be to perform the same evaluation with the target group (the elderly). This will allow us to understand how the user experience differs between both groups, and what considerations should be taken into account when designing games based on voice interfaces for the general population or for older adults.

Participants were students or faculty members from the Computer Science School at the University of Costa Rica, randomly asked to participate in the study. Of those participants, 20 were tasked with the evaluation of the Wizard of Oz prototype, individually. The rest of the participants were in charge of evaluating the Google Home Akinator application.

3.3. Data Collection

The evaluation for the prototype and the Google Home application were done separately.

For the prototype, both games were evaluated together, one following the other. To accomplish this, a Bluetooth speaker was set on a desk inside a computer lab, facing a wall to prevent any possible distractions. The wizard was stationed on the opposite side of the room, where he would be able to listen to the participants' interactions and react accordingly with the pre-implemented options of the speaker, following a predesigned script. The 20 participants were asked to individually enter the lab and were only instructed to follow the instructions the smart speaker gave. These sessions lasted between 5 and 10 min per user. Once done playing, the participant was asked to fill in the UEQ questionnaire with their perceptions of the game.

The Google Assistant application was evaluated on a separate day, with a completely different group of people. For this evaluation, participants were asked to enter an individual cubicle. A single Google Home device was placed on an empty desk on the cubicle. Participants were asked to run the Google Assistant Akinator by saying "Ok Google, let me speak with Akinator". They would then follow the game instructions to play a round or more of the game. Once done, they were asked to fill in the same UEQ questionnaire used for the prototype evaluation.

The answers for both evaluations were inserted in the UEQ analysis and comparing tools. The results obtained can be observed in the following section.

4. Results

UEQ questionnaire measures the user experience for the product being evaluated. Using 26 questions, UEQ covers the attractiveness, usability (efficiency, perspicuity, and dependability), and user experience aspects (stimulation and originality) of the product. Users answer each item with a number from 1 to 7, which UEQ mapped to a value scale ranging from -3 to +3. The analyzing tools calculate the mean of the different values, each associated with an aspect, to obtain an aggregated perception of the product. Mean values between -0.8 and 0.8 represent a neutral evaluation, values greater than 0,8 represent a positive evaluation and values lower than -0,8 represent a negative evaluation. Using the same three-point scale, UEQ can be further grouped into three categories that evaluate the experience quality. These categories being:

- Attractiveness quality: Positive or negative assessment of the appeal of the product.
- Pragmatic quality: Usefulness and usability of the product to achieve given goals.
- Hedonic quality: Product’s ability to stimulate, engage, and appeal to a person’s pleasure.

Table 1 shows the mean value results of the evaluation for the prototype developed and for the Google Home Akinator application, for the attractiveness, usability, and user experience aspects. Figure 2 shows the same results in a graphical representation. Figure 3 groups together these results into the pragmatic and hedonic categories.

Table 1. UEQ scale results for both evaluations, Wizard of Oz and Google Assistant application. The mean value of the answers given by participants, and the variance of the results is given.

Aspect	Wizard of Oz Prototype		Google Assistant Application	
	Mean Result	Variance	Mean Result	Variance
Attractiveness	2.272	0.47	1.561	1.15
Perspicuity	2.355	0.62	1.550	1.45
Efficiency	1.974	0.67	1.150	0.39
Dependability	1.566	0.35	0.950	0.83
Stimulation	1.829	0.83	1.200	0.76
Novelty	1.579	0.74	0.550	0.33

Finally, Figure 4 shows the distribution of answers per item for the prototype evaluation, which means, the percentage of users who chose a given option as their answer. For example, item 1 shows that 60% of the participants answered with a 7, indicating they saw the prototype as enjoyable (+3 in the UEQ scale). Special attention must be given to items 8 (representing the prototype’s predictiveness), 11 (supportiveness), 18 (motivating), and items 2, 4 and 12 (understandable, easy to learn, and clear respectively) which form the perspicuity aspect.

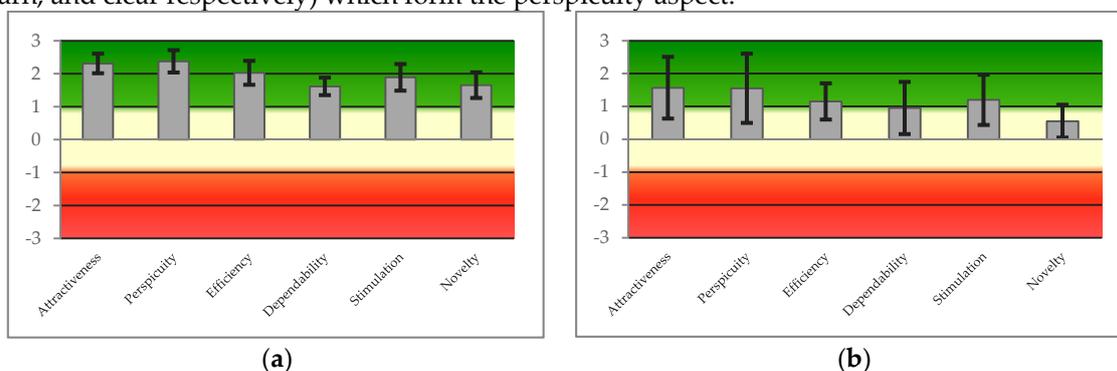


Figure 2. Graphical representation of the results for items evaluated in the UEQ questionnaire: (a) The results for the prototype evaluation, (b) the results for the Google Home application evaluation.

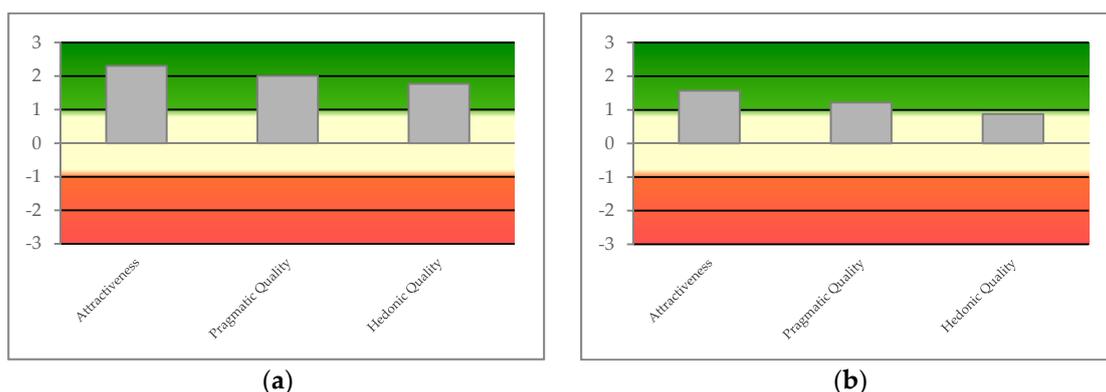


Figure 3. Evaluation results from an attractiveness, pragmatic, and hedonic quality point of view: (a) The results for the prototype evaluation, (b) the results for the Google Home.

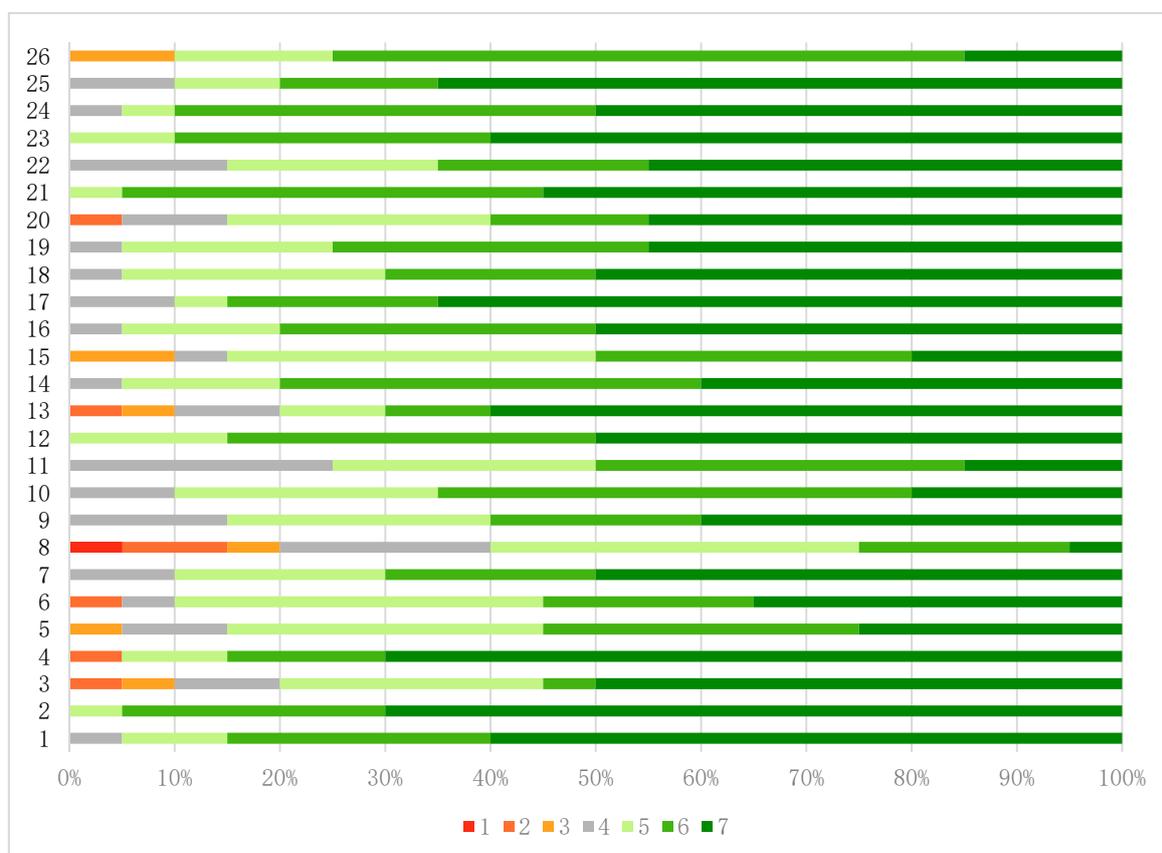


Figure 4. Distribution of answers per item for the prototype evaluation.

5. Discussion

By observing the results shown in Table 1 and Figures 2 and 3, it is simple to see that the prototype got positive user experience results. According to the UEQ scale, an application has positive results in a category if its mean value exceeds 0.8 points. In all the categories evaluated, the prototype presents a pleasant and positive user experience. The values obtained for the perspicuity quality indicate that the application was simple and clear to use, something essential to achieve for games focused on older adults [7–9]. Similarly, the quality of stimulation, an indication of whether a prototype manages to engage with the user and stimulate the user’s cognitive ability, has a result that exceeds 1.8, which is quite positive. This stimulation of the thinking and memory mental processes helps to protect against cognitive decline [1–4,6,8]. Moreover, the high value of the pragmatic quality results, observed in Figure 3a, reinforces the idea that the system is able to entertain and stimulate

the user, while the good hedonic quality indicates that the application manages to capture the attention and engage the user.

Even though these good results were obtained, improvement points can also be identified from the evaluation. Hedonic qualities, shown in Figure 3a, can still be enhanced. Specifically, these results point out that the naturalness of the interactions can be improved. The main suggestion is including additional comments or responses during the gameplay, which would appear to be another person playing against the user. An example of this is including phrases when playing Akinator to point out it is having a hard time guessing a character or is close to doing so. This is expected to improve the general user experience.

When comparing these results with its implemented counterpart, the Wizard of Oz prototype got better user experience results in all aspects. This gives us insight into the fact that, even with the high advances of smart speaker technology, its interaction with a user is still not as natural and intuitive as needed. A human-operated system still makes better work at engaging with a final user. This can be explained by the faster reaction the Wizard of Oz has when responding to a user's input, which produces a more fluent interaction. Additionally, since the Wizard can interpret everything the user says, it has fewer possibilities of answering wrongly to the users' input or prompting for the input again. On the other hand, the Google Assistant application is dependent on a list of intents or possible inputs to which it can respond. This causes more friction in the interaction with the user if the person does not closely say what the application is expecting.

This difference obtained between both evaluations and its explanation hints towards a simple solution to enhance an implemented product. Obtaining a comprehensive list of the different responses and comments that a user can make while playing the games. This will prevent the user from having to repeat his interactions or the application providing an unexpected response. Another improvement would be to increase the randomness of the responses given by the device, as long as it is a correct response to give.

The closest aspect of the UEQ scale between evaluations is the stimulation, with the application having a mean result of 1.2 against the 1.829 of the Wizard of Oz (refer to Table 1). Given that both applications implement a version of Akinator, is logical that they are similar. However, since the prototype also implements the word game, an increase in its stimulation was expected.

If we focus on the answers for the first evaluation observed in Figure 4, the first impression is that the results obtained are quite positive. This indicates that both the proposed idea as well as its implementation were well received. Among the high points of the evaluation are the motivating and supportiveness items, with more than 90% and 70% of answers being positive, respectively. This indicates that the system engages and motivates the participants to use it, while supporting and stimulating their cognitive processes.

Ease of learning, clearness, and understanding aspects, shown in Figure 4, have a very positive answer distribution. This complies with the minimum rules and easy understating principles identified for selecting and implementing the games that were alluring to the aging population. If the same simple interface could be transferred to a fully functional smart speaker, it could have good effects on the user experience when interacting with such applications. Of course, the capacity to achieve this would depend on the existing limitations the smart speaker technology has to date.

Finally, the lowest item present on the answer distribution of Figure 4 is the predictiveness of the system. This can be explained especially for the Word Chains game. When users lost, it would not provide the reasons why they lost, only mentioning that a word they used did not follow the sequence. However, some of the observed reactions from the users point that sometimes they would not agree with this, and thus may think that the system was unpredictable with the losing criteria. By giving the user the option to get a detailed explanation of why the word was not part of the sequence, this item could potentially improve.

6. Conclusions

The idea proposed by this paper, and the voice interface design evaluated, received positive results for its user experience. The pragmatic quality and stimulation aspect results show that games

as the ones prototyped can fulfill the objective of making the player use his cognitive, mental, and thinking processes. Once the same evaluation is done with the elderly, considerations for designing this type of interface for them could be identified, as well as the perception differences between general population users and older adults.

However, it is also drawn from the evaluations that smart speaker technology is not yet able to offer the tools needed to implement a voice-based game with the desired optimal user experience. It has limitations regarding the naturalness of the interfaces. The experience of interacting with such devices still requires an additional cognitive effort not found when interacting with another person.

Finally, punctual recommendations were identified to enhance the user experience accomplished when implementing games with voice interfaces. These being, building a comprehensive list of possible user interactions and various possible responses for each interaction, optional capacity to provide the reason why the user lost a round if it applies to the game, and throwing different expressions that human players would occasionally say in an actual game.

References

1. Bamidis, P.D.; Vivas, A.; Styliadis, C.; Frantzidis, C.; Klados, M.; Schlee, W.; Siountas, A.; Papageorgiou, S.G. A review of physical and cognitive interventions in aging. *Neurosci. Biobehav. Rev.* **2014**, *44*, 206–220, doi:10.1016/j.neubiorev.2014.03.019.
2. Verghese, J.; Lipton, R.; Katz, M.; Hall, C.; Derby, C.; Kuslansky, G.; Ambrose, A.; Sliwinski, M.; Buschke, H. Leisure Activities and the Risk of Dementia in the Elderly. *N. Engl. J. Med.* **2003**, *348*, 2508–2516, doi:10.1056/NEJMoa022252.
3. Goldberg, S.E.; Whittamore, K.H.; Pollock, K.; Harwood, R.H.; Gladman, J. Caring for cognitively impaired older patients in the general hospital. *Int. J. Nurs. Stud.* **2014**, *50*, 1332–1343, doi:10.1016/j.ijnurstu.2014.02.002.
4. Coyle, J. Use It or Lose It—Do Effortful Mental Activities Protect against Dementia? *N. Engl. J. Med.* **2003**, *348*, 2489–2490, doi:10.1056/NEJMp030051.
5. Tertiary Education. Available online: <https://ourworldindata.org/tertiary-education> (accessed on 9 July 2019).
6. Woods, B.; Aguirre, E.; Spector, A.; Orrell, M. Cognitive stimulation to improve cognitive functioning in people with dementia. *Cochrane Database Syst. Rev.* **2012**, doi:10.1056/NEJMp030051.
7. Plaza, I.; Martín, L.; Martín, S.; Medrano, C. Mobile applications in an aging society: Status and trends. *J. Syst. Softw.* **2011**, *84*, 1977–1988, doi:10.1016/j.jss.2011.05.035.
8. Ijsselsteijn, W.; Nap, H.; de Kort, Y.; Poels, K. Digital game design for elderly users. In Proceedings of the 2007 Conference on Future Play, Toronto, ON, Canada, 15–17 November 2007; pp. 17–22.
9. Mubin, O.; Shahid, S.; Al Mahmud, A. Walk 2 win: Towards designing a mobile game for elderly's social engagement. In Proceedings of the 22nd British HCI Group Annual Conference on People and Computers, Liverpool, UK, 1–5 September 2008; Volume 2, pp. 11–14, doi:10.1145/1531826.1531830.
10. Olive, J.P. The voice user interface. In Proceedings of the GLOBCOM 2009, Honolulu, HI, USA, 30 November–4 December 2009; pp. 2051–2055.
11. Spiliotopoulos, D.; Stavropoulou, P.; Kouroupetoglou, G. Spoken Dialogue Interfaces: Integrating Usability. *Hci Usability E-Incl.* **2009**, *5889*, 484–499.
12. Ortiz, C.L. The Road to Natural Conversational Speech Interfaces. *IEEE Internet Comput.* **2014**, *18*, 74–78.
13. Suhm, B. Towards best practices for speech user interface design. In Proceedings of the 8th EUROSPEECH, Geneva, Switzerland, 1–4 September 2003.
14. User Experience Questionnaire, Available Online: www.ueq-online.org (accessed on 8 May 2019).

