

Voice Commands in Home Environment - a Consumer Survey

Hannu Soronen¹, Markku Turunen², Jaakko Hakulinen²

¹ Unit of Human-Centered Technology, Tampere University of Technology, Finland

² Department of Computer Sciences, University of Tampere, Finland

hannu.soronen@tut.fi, markku.turunen@cs.uta.fi, jaakko.hakulinen@cs.uta.fi

Abstract

We studied with telephone survey the opinions and ideas of 1004 Finns concerning domestic technology and controlling it with voice commands. There is distrust towards commanding your home environment with voice; the users doubt especially its functionality. Voice commands are regarded as unpleasant. The most positive conceived aspects of voice commands are the speed of control and the ability to free your hands for something else. Voice feedback, however, is really appreciated. Especially it is welcomed instead of alarm beeps and blinking lights. It is possible that users need first to get accustomed that a device speaks to them. Finally, our studies show major changes in user attitudes when they actually use speech applications.

Index Terms: smart home, speech interface, voice command, consumer survey

1. Introduction

Speech interfaces are moving from telephone-based spoken dialogue systems to ubiquitous computing solutions. Speech provides interface solutions that can be embedded in the environment and included into small mobile devices. In these new uses, speech is usually combined with other modalities. In the project “Ambient Intelligence based on sound, speech and multisensory interaction” (<https://tapla.cs.tut.fi/>), the goal is to develop methods for new type of ubiquitous applications based on sound, speech, machine vision and multimodality. The research covers technology, including audio processing and machine vision, applications, and usability and user interface development. In practise, we are focusing on the use of speech and other modalities in home and similar environments. Currently, we are designing an interface to a home entertainment system featuring speech, keypad and gesture input visual and audio output. Later the work will expand to wider home control functionality. The project works both with the solutions available today and the solutions of the future, that are to be enabled soon when technology matures. Prototype applications that are interesting to consumers and work as part of the environment will be developed.

The new applications areas, and the applications of speech and other new modalities in them, provide numerous challenges, not only for technology, but also to usability and user interface research. Since the new applications bring new kinds of use context and use style, the attitudes and expectations people have to these systems are currently unknown. Working with home environment is particularly challenging, since attitudes towards home tend to be very strong. Some people like to keep their home as a peaceful place of relaxation, while others are keen to try out any new technology to improve their home. There is need to know

more about user attitudes towards the use of speech and other modalities in these new usage scenarios.

2. The Research

We realized that our research teams’ views on the speech interface in home environment were based on previous studies with just small samples and consisting mainly of qualitative data. There was no way to make generalizations based on that data. To get a clear idea of overall acceptance of speech interface at homes, a quantitative data was badly needed. We planned a questionnaire consisting of 82 research questions and Likert-scaled opinion claims which would measure:

- what functions at home environment people would or would definitely not like to control with speech
- what are the reasons for acceptance/rejection of the speech interface at home.

By conducting the survey we would get crucial information of the future users’ focused interest towards the idea of speech interface at home and therefore the potential of the consumer market and specific consumer segments in Finland. This way we would be able to concentrate our future work on the most desired use context and functions.

2.1. Survey

A detailed telephone survey was conducted for a total 1004 interviewees in November and December 2007. The interviews were performed by Taloustutkimus, a Finnish market research company. The random sample was allocated both geographically and for the age distribution (18-64 yrs), and is representative of the whole Finnish population. Both sexes were represented equally. The interview lasted ca. 15 minutes and it contained only questions on this subject.

2.1.1. *The acceptance of voice commands at home*

The overall charm of speech interface seemed to be very little. When asked “how interested would you be in controlling your home with following means” on a four-point Likert scale, speech (mean 2.1; sd 1.0) was considered less appealing than touch screen (2.5; 0.9), remote control (2.4; 0.9), computer (2.4; 1.0) or mobile phone (2.3; 0.9). Only digital television was slightly less appealing, probably due to its poor introduction to Finnish viewers. The result was somewhat surprising considering that speech is usually perceived as an easy and natural way of interacting with the environment. It appears that interacting with technology might be another matter.

Next we asked about certain functions at home and their suitability for speech control. The functions were obtained from a range of partners in the project. These results can be seen in Table 1.

Table 1. Users' interest in controlling home with voice commands

N=994-1004 (Scale 1= "Not interested at all" – 4= "very interested")	Mean	SD
Warming up the car	2,4	1,2
Controlling the lights	2,3	1,1
Dictating and sending email	2,3	1,1
Making a phonecall	2,2	1,1
Recording tv programme with timing	2,2	1,1
Choosing and listening to music	2,2	1,0
Searching for information on the Internet	2,1	1,1
Opening the front door for oneself	2,1	1,1
Controlling the room temperature	2,1	1,0
Filling an electronic form for authorities	2,1	1,1
Watching TV (changing channel, adjusting volume, shutting down)	2,1	1,1
Ordering for domestic help by dictating	2,1	1,1
Watching digital photos from TV or computer	2,1	1,0
Locking the doors	2,0	1,1
Heating the sauna	2,0	1,1
Replacing the keyboard and mouse with voice commands	2,0	1,0
Opening the door after hearing doorbell ring	1,8	1,0
Using the blinds	1,8	1,0
Using the kitchen stove	1,8	1,0
Flushing the toilet seat	1,7	1,0
Playing a video game with voice commands	1,4	0,7

Apparently controlling your home environment with voice does not seem very appealing to Finnish users. Even the most helpful functions, as opening the front door when you are arriving at home with hands full of bags, are not warmly welcomed. All the suggested functions received moderate or strong verdict from the interviewees. Based on this data and combining it with more qualitative impressions from previous studies we concluded that controlling the home with voice appears to Finnish users as one whole. It is the very idea of giving voice commands that is not accepted at this point, even when controlling with speech would be very handy or helpful. There seems to be a relatively high threshold for using voice to control technology. What could be behind these impressions?

2.1.2. The reasons for acceptance/rejection

We had several questions for determining the reasons behind users' acceptance or rejection. A majority of Finns (61%) feel that the idea of speaking to a device is unpleasant, 37% agree totally with this claim. 63% think that voice commands do not make a device more pleasant to use. Even more consumers (67%) feel that controlling with voice is unreliable since the system may misinterpret the given command. Winning the trust of the consumers may thus require considerable amount of work in the field of offering positive examples of voice control and the additional value it may bring in certain contexts. Finnish consumers lack the experience in speaking to a device even though a number of them carry a device perfectly capable of this kind of interaction: the mobile phone.

To make an even deeper analysis of the reasons for rejection we analyzed those 95 respondents who answered that they would not pay anything for speech controlled domestic technology either as a one time payment or a monthly fee. They rejected the whole idea mainly because they feel the system is utterly useless, they had no need for it (63%). Further, 21% felt that the system is unreliable, it would not work. 14% of them would not want more technology in their homes. High costs were considered a reason for rejection in 6% of the cases. Other single reasons mentioned were fear of becoming too passive, and the traditional appreciation of doing things with your own hands. Multiple answers were allowed in this particular question, so the sum of percentages is over 100.

On the positive side, consumers seem to think that controlling technology with voice commands is faster than by pressing buttons or clicking with mouse (63% agreed with this opinion). The idea of smoothness and speed becomes more appealing for example with home media center: instead of making several time consuming choices from the menu one could command the system to play certain song or album, even from another room. Another accepted motivation for using voice commands, is freeing your hands. You can easily continue your routines and still are able to control the chosen device, for example the TV. 54% of the respondents agreed that this would be pleasant. Still, 45% felt that it would be pleasant to get rid of all the remote controls if voice command would take the charge instead. With the introduction of digital set-top-boxes the living room table has once again received another unwelcome device. According to the discussion in the public media, frustration seems quite high towards these ugly and cumbersome devices.

2.1.3. Voice feedback

Interaction with technology is of course two-way. Therefore we also examined the users' acceptance of getting voice feedback in specific contexts. This time the reactions were much more positive and accepting. Some of these results are presented in table 2.

Spoken warnings or interpreted alarms would be very appealing to Finnish users instead of ambiguous beeps and blinking led lights which you have to interpret yourself with the multipage aid from the manual. Somehow it seems far more acceptable and "natural" that the device speaks than humans speak to the device. It is our view that his approach should be appreciated when introducing new technology to home environment.

Table 2. Users' interest in voice feedback in home environment

N=993-1004 (Scale 1="Not interested at all" – 4= "very interested")	Mean	SD
Fire alarm	3,5	0,9
Alarm for kitchen stove still on	3,4	1,0
Alarm for water leakage	3,1	1,0
Safety camera alarm	2,8	1,1
Reminding of medication	2,8	1,1
Alarm for open doors or windows	2,6	1,1

Statement	Interested in voice commands		Mean (Likert scale 1-4, disagree-agree)	SD	Sig.
	very little	N			
It is important that home is equipped with the latest entertainment electronics.	very little	112	1,6	0,8	***
	very much	125	2,4	1,0	
It is important that home is equipped with fast internet connection.	very little	111	2,3	1,1	***
	very much	125	3,1	1,0	
Technology makes daily routines easier.	very little	112	3,0	0,8	***
	very much	123	3,6	0,5	
Technology improves enjoyment and pastime at home	very little	112	2,7	0,9	***
	very much	125	3,3	0,7	
I don't want my home to get too technical	very little	112	3,1	1,1	***
	very much	125	2,5	1,0	
I get joy and pleasure from purchasing new technology	very little	111	2,0	0,9	***
	very much	124	2,7	1,0	
It is important to be able to fix and repair one's home by him/herself	very little	110	3,3	0,8	-
	very much	125	3,3	0,8	
Automatic system can well take care of airconditioning, lighting, home appliances and heating.	very little	111	2,4	1,1	***
	very much	123	3,2	0,8	
It is important to be economical when purchasing technology for home	very little	112	3,3	0,9	-
	very much	125	3,1	0,9	
I'm afraid that machines start to trouble my life.	very little	112	1,8	1,0	-
	very much	124	1,8	0,9	
New domestic technology is pleasant to use.	very little	111	2,6	0,9	***
	very much	125	3,2	0,8	
I think that using technology is addictive.	very little	112	2,2	1,1	-
	very much	123	2,3	1,0	
I'm afraid that my computer gets infected with a virus which will destroy personally valuable information.	very little	107	1,9	1,0	*
	very much	125	2,2	1,0	
I'm worried that a great deal of people will lag behind others in the technical development	very little	112	2,0	1,0	***
	very much	122	2,5	0,9	
Implementation of new technology is easy for me.	very little	112	2,5	1,1	**
	very much	124	2,9	0,9	

Statistical significances, independent samples t-test: - = no significance, * = p<.05, ** = p<.01, *** = p<.001

Table 3. User attitudes and values in connection to acceptance of voice command in home environment

2.1.4. Identifying the consumers

The next task was to analyze the data and to identify those who were well disposed towards voice commands. As we soon found out, the uniform rejection of the idea of speaking to a device also meant that specific user segmentation based on socio-economic background data did not work. There were no statistically significant differences between the sexes, level of education, profession, number of household members, computer skills, income level or previous experience with internet. We were left almost empty handed. Only age seemed to explain some of the differences: the older generation expectedly showing less interest. Still we had placed several attitude and value oriented questions in the questionnaire, so the next task was to crosstabulate these factors with the voice command opinions. The results were clear and to large extent statistically significant as can be seen from Table 3. The overall acceptance of speech interaction with technology is strongly value based. Those who regard voice commands for their homes more positively also seem to get more pleasure from using technology in general than those who reject. They

are also more trustful, and less worried about the negative outcomes of technological development. Based on these results it is possible to screen out the customer types and this will be the topic on next phase of the research.

3. Discussion and conclusions

We studied with telephone survey the opinions and ideas of 1004 Finns concerning domestic technology and different ways to control them. Our specific target was consumers' interest in voice commands in general and also in certain defined contexts at homes. According to our results there seems to be distrust towards commanding home environment with voice, especially its functionality. Voice commands are regarded as unpleasant, which tells us of a threshold to speak to a technological device. Even though Finns are known for their technology-mindedness, they find it odd to speak to "a system". This is no doubt in some connection to the idea of projecting human-like features in the system that communicating with speech may elicit. Users easily project mental states, life-like essences and social rapport even with electronic toys such as Aibo [2]. There can be other culturally mediated preferences too: in a recent study it was found out that the Japanese thought the TV-interface was better with a remote control whereas US participants preferred it with voice control [3].

The results show that the most positive conceived aspects of voice commands are the speed of control and the ability to free your hands for something else. In the design phase of our project we will pay special attention to these findings. We will focus on the functions that will let the user control domestic devices when they are already doing something and would not want to disrupt it [1]; and to contexts where the user wants to control several subfunctions of the system at the same time, for example setting the media center in a certain mode which would acquire numerous sequential choices from the remote control. Our results also revealed that voice feedback, on the contrary, is really appreciated. Especially it is welcomed instead of alarm beeps and blinking lights. It is possible, we argue, that before Finnish users get used to the idea of speaking to technology, they need to get accustomed that a device speaks to them. After this change the readiness to make a speech contact with technology in home environment could emerge.

3.1. User Experiences

In order to find out how user expectations may change, we studied user acceptance of speech applications in another Finnish research project (<http://pums.fi/>) by comparing user expectations and experiences (perceptions) before and after they used a prototype of a mobile speech application. The application in question is TravelMan [4], a multimodal mobile application providing route guidance for public transport in Finland, such as metro, tram, and bus traffic in cities and long-distance traffic in the rest of the country. There are two main functions, planning a journey and interactive guidance during the journey. In the journey planning phase, a user enters the departure and destination addresses or locations using speech input or predictive text input. The user interface includes a tight integration of graphics and speech outputs. This means, that everything appearing on the mobile phone screen is also read out loud by the speech synthesizer. In mobile applications, speech outputs are considered particularly useful for visually-impaired users. Still, they can be helpful to all users, since in the mobile context of use we all have sometimes limited vision (e.g., due to need to monitor our surroundings). There are speech synthesizers in some new mobile phones, but little is known their usefulness, usability, and user acceptance. Furthermore, there has been always the question of acceptance of synthesized speech output in general.

The TravelMan application has been in public pilot use since spring 2007. The application had over 1000 real users in the pilot phase, and it was well received by the users. In order to find out the efficiency and user experience of the application, we arranged user evaluations in a lab environment. 38 students from the local university participated in the evaluation (27 male, 11 female). Their age ranged from 18 to 45 years. Both objective and subjective metrics were collected using a service quality metric [5], to analyze the interactions and elicit feedback from the participants. For more information, see [6].

To summarize the results, synthesized speech outputs, which represent the state-of-the-art in current mobile synthesizers available for Finnish, were received rather poorly in overall subjective opinions, especially when future use is considered. Otherwise, the participants rated synthesized speech outputs to be fast, clear, easy to use, and error free, but because of lack of naturalness and pleasantness, they ranked it low in overall scores, both absolutely and especially when

compared to expectations which were quite high similar to the consumer survey.

For speech inputs, objective results show that speech is both in theory and practice the most efficient input method, even with relatively high error rates and slow response times. Furthermore, users also rated speech input very high compared to their expectations, so our results show that when people use speech applications, their low expectations may change, even with non-perfect speech recognition, like in our case study (the speech recognition accuracy was 70%, resembling realistic results, not 99% accuracies seen in advertisements). However, as shown in the consumer study, people do not have too high expectations for speech input, and although user experiences are good compared to expectations, other modalities easily outperform speech in subjective ratings, like in the case of TravelMan.

It is also interesting that we could not find correlation between the actual recognition accuracy and the perceived robustness of speech inputs. Neither was it explaining usefulness ratings of speech inputs or their future use.

Finally, it is interesting that users expected speech outputs to be more useful than speech recognition, but perceived its quality to be less, both in relative and absolute sense, after they used application which included both. This is in contrast with the results of the consumer studies presented here.

Ecological validity of the evaluation conditions is worth considering as well. In comparisons of laboratory experiments and real usage of speech applications great differences has been found in previous studies. In mobile situations, the hands and eyes free interaction may favor speech and limit the usefulness of other methods. For example, pen-based soft-key text input has been shown to be slower while walking. Furthermore, in our laboratory conditions, speech synthesis and inputs did not provided any added value. However, as seen from the consumer survey, people expect quite a lot from speech outputs.

4. Acknowledgements

This work was supported by the Technology Development Agency of Finland (TEKES) under the Ubicom-programme in the "Ambient Intelligence Based on Sound, Speech and Multisensor Interaction"-project (TÄPLÄ, grant 40223/07).

5. References

- [1] Bradbury, J.S., Shell J.S., Knowles C.B. Hands On Cooking: Towards an Attentive Kitchen. CHI 2003: New Horizons, April 5-10, 2003.
- [2] Friedman, B., Kahn P.H., Jr. Hagman Hardware Companions? – What Online AIBO Discussion Forums Reveal about the Human-Robotic Relationship. CHI Letters. Volume No.5, Issue No.1. CHI 2003. April 5-10 2003.
- [3] Tan, G., Takechi M, Brave S, Nass C. Effects of Voice vs. Remote on U.S. and Japanese User Satisfaction With Interactive HDTV Systems. New Horizons CHI 2003, April 5-10, 2003.
- [4] Turunen, M., Hakulinen, J., Kainulainen, A., Melto, A., Hurtig, T. Design of a Rich Multimodal Interface for Mobile Spoken Route Guidance. In Proceedings of Interspeech 2007 - Eurospeech: 2193-2196, 2007.
- [5] Parasuraman, A., Zeithaml, V.A. and Berry, L.L., "SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality", *Journal of Retailing*, 64, 1, 1988.
- [6] Melto, A., Turunen, M., Hakulinen, J., Kainulainen, A., Heimonen, T. A Comparison of Input Entry Rates in a Multimodal Mobile Application. In Proceedings of Interspeech 2008.