

Telephone based speech interfaces in the developing world, from the perspective of human-human communication

Soogandhree Naidoo (snaidoo2@csir.co.za, +2712 841 2876) and Etienne Barnard (ebarnard@csir.co.za) CSIR / University of Pretoria Human Language Technologies Research group

Abstract

Telephone-based information services hold tremendous potential as a means of information access in the developing world [5]. However, surprisingly little research has been done on the design of such systems – whether basic Interactive Voice Response systems, or more sophisticated systems using speech recognition and speech synthesis. We report on theoretical work, based on research by C. Nass [1], as well as experiments aimed at addressing this situation. We have found that the design of telephone-based services can benefit significantly from detailed consideration of the characteristics of human-human communication. This appears to be particularly true in the developing world.

Summary

Interactive Voice Response systems (IVRs) are an ideal platform for delivering information to users in the developing world. However, such systems were initially designed for technologically sophisticated users in the developed world. It is therefore an important challenge to design telephony interfaces for users with low literacy and limited exposure to technology. In addition, IVR technologies such as synthetic speech and speech recognition technologies are still fairly new in the developing world, and are not widely used. In this work, we propose that in a developing country - where a substantial fraction of users are not technologically sophisticated - IVR systems should be designed to mimic real-life conversations. These conclusions are based on the theoretical framework developed by Professor Clifford Nass at Stanford University, and on observations of user interaction during experiments testing the usability of IVR systems..

These experiments were carried out in a semi rural area and most users were unfamiliar with technology.[4][5] The experiments were performed with the aim of focusing on user performance with regard to IVR telephone-based systems, and to better understand the needs and requirements of the typical South African. The experiments were carried out using two test groups. One interacted with the system in their home language and the other in English. User performance was documented and their reactions and behavior recorded. From observations it was noted that even when the system was explained to users and they were told that they were talking to a computer, their reactions and behavior were strongly influenced by human-human interaction. When asked to vocally input information, some individuals would “chat” to the system. Individuals would also question the system (something that the system is unable to handle). When experiencing difficulties hearing the voice prompts, individuals would ask the system to speak louder.

The experimental observer was not at any time asked to increase the volume, as one would expect from users with technological experience. The behavior displayed by the observed users of the IVR system can be explained by Nass's suggestion of looking at speech interfaces from an evolutionary perspective.[1]

Until recently, before computers systems were able to synthesize or recognize speech, speech was a capability unique to humans. The human brain has developed to differentiate between human speech and other audio occurrences. Therefore, the slowly-evolving human brain reacts in certain ways to voice stimuli, and has certain expectations regarding communication by voice. Nass affirms that the human brain operates using the same mechanisms when interacting with speech interfaces as when conversing with other people.[3] The same principles that guide human –human interaction seem to apply to the interaction between human and voice interfaces.

The human brain has developed to differentiate between human speech and other audio occurrences. Until recently, when speech was heard the automatic conclusion was that a human had spoken. Nass goes on to state, "Individuals behave towards and make attributes about voice systems using the same rules and heuristics that they would normally apply to humans" [1]. With the advent of speech synthesis, speech no longer equates to a human speaking. However the same social rules and expectations regarding spoken interaction still seem to apply. To enable usable speech interfaces these social rules need to be taken into consideration when designing IVR systems.

Given the requirements of typical users in a developing country, like South Africa, interfaces must be designed to encourage usability and error avoidance. Users, who lack technological sophistication, are easily intimidated by "new technology". Users, who do not fully comprehend the limitations of an IVR system, expect the same socially acceptable behavior from interaction with such systems. Understanding how and more importantly why users interact in a certain way will aid in building user-friendly systems. It is essential to keep in mind how people react during social communication when designing voice interfaces. Nass identified ten fundamental principles with respect to processing speech, producing speech, and spoken dialogs for humans from the perspective of evolutionary psychology and demonstrated that these principles apply to human users' interaction with speech interfaces.[2] However the questions still remains: how should these principles be adapted and applied in a developing country? We are producing a list of guidelines, based on Nass's work, regarding the design and development of speech-interfaces for users within a developing country.

References:

1. C Nass, L Gong, (2000), "Speech interfaces from an evolutionary perspective: Social psychological research and design implications", Communications of the ACM, Volume 43, Issue 9 (36 – 43)

2. Nass, C. & Gong, L. (in press). Ten principles for designing human-computer dialogue systems: A perspective from evolutionary psychology. In D. A. Dahl (Ed.), Practical spoken dialog systems. Kluwer Academic Publishers.
3. Nass, C. & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81-103.
4. Easton H, Nakene M, Naidoo S and Kafidi PL, "Validating the user-centred design process within a developing context", CHI-SA 2003 Conference, Johannesburg, 2003.
5. Barnard E., Cloete L., Patel H, "Language and Technology Literacy Barriers to Accessing Government Services", LECTURE NOTES IN COMPUTER SCIENCE, Vol. 2739, 2003, Pages 37-42

Contact Details:

Soogandhree Naidoo: 012 841 2876