

Relationship between Affordance and Cultural Conventions in the Design of IVR Systems for Oral Users

Tembaletu J. Ndwe, Nomusa Dlodlo

Abstract—This paper addresses the significance of affordance in association with cultural conventions in the design of voice user interfaces (VUI) for orally grounded users in the global South (i.e., developing countries of the world). The paper demonstrates that sensitivity towards cultural conventions which subsequently bring about affordance has more credence than the objective usability measures of effectiveness and efficiency as defined by the International Standards Organization (ISO). This demonstration is done with the aid of two case studies of Interactive Voice Response (IVR) systems that were developed for users in developing countries of Southern Africa. The paper specifically presents the concept of taking into consideration the shift between the standard cultural norm and the situational norm experienced during the usage of the IVR system. We have established that orally grounded technology users prefer a VUI that allows them to transfer easily from their standard cultural norms to the situational norm that is determined by the context of use of the presented technology.

Index Terms— Cultural conventions, IVR systems, Oral users, affordance, usability.

I. INTRODUCTION

IT is anticipated and expected that speech technology, through the use of the normal telephone, will have a substantial impact in the realization information access and delivery of information services in the developing nations of the world [2]. This is due to the fact that cellular technology penetration is growing more rapidly in the developing countries even more than the developed world [10]; [3]. Another important reason to take into account is the fact that the majority of the people living in the developing world have a strong oral tradition with high levels of illiteracy. This phenomenon is even more important to consider within the realm of the region of Southern Africa where a strong oral tradition exists amongst a large low literacy population [1].

Manuscript received November 11, 2013. This work was supported in part by the Rhodes University Center of Excellence in Distributed Multimedia, Meraka Institute-CSIR, and the National Research Foundation (NRF).

Tembaletu Jama Ndwe is with the Rhodes University's Department of Computer Science, P.O. Box 94, Grahamstown, 6140, South Africa phone: 046-603 7478; fax: 046-636 1915; e-mail: j.ndwe@ru.ac.za

Nomusa Dlodlo is with Meraka Institute-CSIR, P.O. Box 395, Pretoria, 0001, South Africa; e-mail: NDlodlo@csir.co.za

Speech technology is a feasible channel for information access in low-literacy individuals since listening and talking do not require literacy. In this paper we support the use of speech technology as a means of information access and present an essential attribute that needs to be incorporated in the design of speech technology for orally grounded users in the developing regions of Southern Africa.

The present research study utilizes two case studies of the development of two IVR systems namely the OpenPhone and the Beautiful Game Results (BGR) systems. The OpenPhone system is an IVR system that allows its users to access information on caregiving for HIV/AIDS infected children in Botswana, Southern Africa. The BGR system is a fun system that allows its users to access the results of recently played soccer matches in the Professional Soccer League (PSL) of the Republic of South Africa. The two IVR systems are contrasting in their manner of use as the OpenPhone system has a serious context about a disease that is pandemic in the region of Southern Africa and either infects or affects everybody in the region. On the contrary, the BGR system is a cheerful application that allows its users to interact with a telephonic access to results of soccer games in the Republic of South Africa, where soccer is the favorite sport [16]. The contrast between the two IVR systems has allowed the researchers to discover oral user attitudes and characteristics that would not have been possible with applications that have a similar context of use. We, the researchers, have achieved the contrast and comparison between the two IVR systems by running usability experiments of comparing two modalities of interaction that were presented to recruited intended users for each of the systems.

The two modalities are the speech-enabled modality and the Dual Tone Multi Frequency (DTMF) which are different modes of interaction with an IVR system. In the speech-enabled interaction mode the users interact by giving verbal commands to the system and the system responds by giving back verbal information. The speech-enabled mode is similar to a normal conversation to another person. On the other hand, the DTMF system allows the users to interact with the system by using the telephone keypad as the sole input device. The DTMF system output presents a menu that instructs the user to press a particular number on the phone keypad that corresponds to a particular effect.

For example: *System: To learn about common sicknesses press 4, to learn about ARV medication press 5.*

The user then reacts by pressing whichever number that corresponds to the task that they want to carry out.

The researchers are not aware of any other research or publication that investigates the relationships between affordances and cultural conventions within the usage of IVR systems in developing countries.

II. AFFORDANCE AND CULTURAL CONVENTIONS

The term of affordance was first coined by Gibson who defined the concept of affordance as:

An actionable attribute of an interaction design between the world and an actor (actor refers to a person or an animal) [7].

Gibson's view on affordance implies a relationship between the actor and the attribute of an interaction design and affordance exists naturally and does not have to be seen, known or desirable. Cultural conventions are constraints or regulations shared by a cultural group [13]. The concept of cultural conventions is demonstrated vividly through the contrast in the two case study IVR systems that are presented in this research as follows:

In operation, the DTMF modality necessitates the user to interact by pressing buttons on the phone. The interaction is muted to a nearby person and therefore, except for the user, no other person would be able to comprehend the content and the context of the interaction between the user and the system. Conversely, in the speech-enabled modality, the interaction necessitates the user to respond by using their voice as in a normal conversation with another human being. This means that a close by person would be able to overhear the content of the interaction, whether unintentionally or otherwise. Other people who have used the same system would certainly know both the content and the context of the interaction as well.

Within the culture of the indigenous people of Botswana, speaking out loudly on matters that are affiliated with sex is an inappropriate practice. It is well known that HIV/AIDS is an illness that is mostly spread through sex, and therefore speaking out loud about HIV/AIDS is not common unless it is through educational media. The stigma about HIV/AIDS is a complex situation that is interlinked with other stigmas including ethnicity, religion, gender and others [15]; [4]. Stigma is considered as one of the main impediments in controlling the disease. It is important to note that almost all of the OpenPhone users are women who are mothers and caregivers to the HIV/AIDS infected children, and subsequently they are HIV/AIDS infected themselves. Being an HIV infected woman, within the socio-cultural ideals of the targeted users bears even more stigmatization, suffering in silence and shame [6]. The targeted users' socio-cultural ideals in combination with the lack of seclusion in the modus operandi of the speech-enabled modality support the postulation that the caregivers' choice of the DTMF modality is based on the privacy of the modality when compared to its

counterpart. The researchers view the privacy issue as the most compelling reason behind the users' choice.

An inherent problem with the stigma of HIV/AIDS is that the users would largely not say that the stigma is the reason for their choice of interaction modality. This is partly because from the lectures that they attend and from the media of radio and television the effects and negative consequences of stigma have been addressed over and over, and caregivers know very well that they are not supposed to yield to stigma. Nevertheless, the HIV/AIDS stigma continues to be a problem in their communities and by nature of stigma, it is something no one wants to talk about or be associated with. This discourages the caregivers in revealing stigma as an influence to their choice of interaction modality. In all the 22 participants who performed both modalities and were asked to compare the 2 modalities, only one participant mentioned that they prefer the DTMF because of its confidentiality, stating that in the speech-enabled system other people can hear what the user says. Interestingly, this particular user got her DTMF tasks incorrect and got her speech-enabled tasks correct during the usability tests and still chose DTMF as her modality of choice.

In contrast to the OpenPhone system's access to a stigmatized issue, the BGR provides access to an issue that is not only free of stigma but something that has a culture of being publicly debated and spoken about loudly in public places, private homes, work, and generally everywhere. It is a normal practice to find soccer fans gathered at work, in public places or over a phone call and debating issues about recent games and such conversations are typically accompanied with exuberance and loud speaking and laughter of fans teasing the supporters of teams that recently lost a game. Soccer fans love to speak out loudly about their favorite teams and this loud and exuberant expression of love for soccer is part of South African soccer culture and it is compatible with the oral tradition of the indigenous people of the country as it promotes expression of opinions and chatting freely on a subject that is well-known and loved by the soccer fans. The exuberance of speaking loudly on soccer issues was evidenced on the BGR whereby participants were speaking loud and fast as they gave commands to the system, something they were not requested to do, as they were only requested to speak clearly. This was in direct contrast to the reserved, monotonous and slow style of interaction that was experienced with the OpenPhone participants. According to studies that were conducted in human emotional sensitivity, loud and fast voice was found to correlate to happiness, joy and confidence whilst a low and slow voice was linked to boredom, grief and sadness [17]; [5]. The researchers believe that the reason for the overwhelming choice of the speech-enabled modality in BGR is because it allows the users to interact with the system in the same manner that they normally address soccer issues which is, simply, loud. Everything about the game in the region is loud; from the vociferous vuvuzela - a plastic horn played in and outside soccer matches that produces a loud monotone note, to the

colorful apparel worn by the spectators, inside and outside the stadia. Likewise, the choice of the DTMF system in OpenPhone was due to fact that DTMF allows the users to exercise confidentiality in an interaction with a stigmatized subject matter within a society that imposes even more stigmatization towards infected women as they are looked upon more negatively than infected males.

In summary it can be stated that during the interaction with OpenPhone system, the users of the technology have a situational norm or a cultural convention of quietness and being mute due to the given situation of interacting with a stigmatized subject. Contrastingly, the BGR users have a cultural convention or situational norm of loudness during the interaction with the technology due to the situation of interacting with a cheerful application that is associated with loud situations.

III. DEVELOPMENT METHODOLOGY AND EXPERIMENTS

In this section the methodology that was used in the development of the two IVR systems and their associated prototypes (i.e., DTMF and speech-enabled modalities) is discussed together with the measurements that were taken during the usability tests and the experimental results that were obtained. The types of measurements that are relevant to this study are stated and the researchers also explain the motivation behind the measurements used.

A. Methodology

The methodology of User Centered Design (UCD) was deployed in the development of the two IVR systems. The choice for UCD was motivated by the researchers' desire to engage the intended users of the technology in every phase of the development lifecycle of the IVR prototype systems. UCD is a methodology for designing technology products/services such that they meet usability objectives. In other words, UCD is a set of activities that need to be taken so as to ensure usability in an ICT product/service design/development. The UCD methodology is a participatory design methodology as it involves the anticipated users throughout the development of the technology. One of the essential activities in UCD is usability testing which allows the researchers, in alliance with the intended users, to evaluate the prototype designs against the technology design requirements.

According to the ISO the definition of usability is:

Usability is the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use [9].

Additionally the ISO standard describes the implied usability characteristics of effectiveness, efficiency, and satisfaction in Table 1.

The ISO 9241-11 standard definition of usability is becoming the main reference of usability [12]; hence it has been used extensively in this research.

USABILITY CHARACTERISTIC	DEFINITION
Effectiveness	Accuracy and completeness with which users achieve specified goals
Efficiency	Resources expended in relation to the accuracy and completeness with which users achieve goals
Satisfaction	Freedom from discomfort, and positive attitudes towards the use of the product

Table 1: Definitions of usability characteristics [9]

B. Usability Experiments Measures

There are typically two types of results that usability experimenters are interested in, which are the qualitative and quantitative results. The objective measurements are quantitative factors that are obtained from empirical measurements and the subjective results are qualitative representations of the intended users' sentiments and satisfaction levels towards the prototypes. According to the ISO definition of usability, as indicated in Table 1, the usability characteristics of effectiveness and efficiency are the quantitative objective measures and the usability characteristic of satisfaction is the qualitative subjective measure of usability. In both IVR systems the subjective results are used and viewed as more dependable than objective results given the exploratory nature of the experiments. Human experience is generally subjective [11], and user subjective reactions to speech interfaces may well be a more important predictor of real world success [8]. Shackel also regards user satisfaction or users' attitude towards the system as the most important aspect of usability [18], and objective measures of system behavior may not suffice in predicting system acceptability [19]. As the study is centered on the first iteration of the IVR systems development, the researchers are more interested in evaluating the technology's feasibility to add value in the lives of the intended users through the measurement of the users' acceptance of the technology since the technology is new to the intended users. The success or failure of the technology depends on the human experience and the study mainly relies on the intended users' view on whether the technology is enough to serve its purpose and meet the user needs and expectations or not. Nevertheless, the study utilizes both the subjective and objective results in unison and regards both as complementary. In the analyses of the results, a good conclusion is considered as one that has support of both types of results.

C. Experimental Results

As mentioned in the earlier paragraph, this research views subjective results as more prominent than objective results even though reference is made to the qualitative objective results. The following subsections discuss the results that were obtained from the usability tests for both IVR systems.

OpenPhone Subjective Results

After completing both sessions (i.e., one in DTMF and another in speech-enabled mode), participants were asked a few demographic questions which requested their age, level of education, home language, and others. They were also asked for their subjective evaluation of the systems and their choice of interaction modality between the DTMF and the speech-enabled modalities.

Out of the 22 participants who interacted with both OpenPhone prototypes (i.e., the DTMF and the speech-enabled prototypes), 13 preferred DTMF, four preferred the speech-enabled system, and five were equally happy with both systems which produced a 59.1% preference of DTMF over 18.2% for speech-enabled system, with 22.7% of users who were undecided. Two people out of the four who preferred the speech-enabled system said that they actually like both but they were choosing the speech-enabled system because they think that it would be a better system for elderly users who might have difficulty using the DTMF system. Another participant mentioned that she preferred the speech-enabled system because she envisions doing something else with her hands whilst interacting with the system. The participants who did not have any particular preference over any of the systems claimed that the two interfaces were the same to them.

The substantial majority of 13 participants who chose the DTMF system had well-defined reasons including perceived faster speed. The most common reason was the ease of use and the fact that the DTMF was easy to follow because, “it is impossible to get lost as the system [DTMF] just tells you what to do.” Another participant remarked that in the speech-enabled system she had to pay more attention because if a command is missed or misunderstood or misinterpreted, then that can be problematic but the DTMF is more “straightforward.” Throughout the interactions, the participants were noted to speak in a low, slow paced and monotonous voice and their facial expressions did not show any particular emotion.

BGR Subjective Results

The BGR subjective questions were conducted in the same manner as in the OpenPhone system by asking similar questions and then asking the participants to compare the two modalities after having performed tasks in both modalities.

In total, out of the entire set of 27 participants who participated in the BGR tests, 23 preferred speech-enabled system and four preferred DTMF, which produced an 85.2% preference of speech-enabled system over 14.8% for DTMF. The BGR users were all decisive in their preferences and none of them preferred both modalities as in the case of OpenPhone.

The 23 users who preferred speech-enabled system had different reasons which included the fact that it presented a more natural way of interaction through speaking rather than the extra effort of pressing numbers. These users considered the speech-enabled system to be easier to use and they felt that

they had more control when giving verbal commands than pressing numbers in the DTMF modality.

The users also perceived the speech-enabled system to be faster even though objective results illustrated that, on average, DTMF outperformed speech-enabled system in terms of the time taken to complete tasks although task completion times using the two systems were not significantly different. Another reason for speech-enabled system preference was novelty. Users mentioned that speech-enabled system is trendy and it is an advanced technology that is beyond their familiarity with DTMF systems that they have used before.

Overall, it was evident how the BGR users were excited in using the BGR system by the way that they conducted the tests. An obvious observation was that users were speaking out confidently, loud and swiftly in positively articulated utterances during the interaction with the BGR speech-enabled prototype. The caller loudness and confident speaking contributed to the very high recognition levels that were experienced in the speech-enabled prototype, “as confident speakers enunciate more clearly, making the recognition task much easier” (E. Barnard, January 10, 2011, e-mail message to author). This was particularly helpful for users who were calling from cellphones in environments that were not ideal for speech recognition purposes. The researchers did not request the callers to raise their voice at any time but the callers selected this style of interaction at their own discretion.

IV. OBSERVATIONS AND ANALYSES

In our experiments for both the OpenPhone and BGR systems we have observed that the choice between DTMF and speech-enabled system is not influenced by the performance levels of the users, which implies that the effectiveness, of the system, as defined by ISO, does not influence users’ choice. This was evidenced by the fact that in both case studies the users preferred systems where they performed equally or less successfully. Also, in the BGR study, the users chose the system that was marginally less efficient to them in terms of time taken to complete the required tasks, which implies that their choice is not influenced by efficiency of the system. From this it can be inferred that the objective usability measures of effectiveness and efficiency, as defined by ISO, do not have a conclusive bearing on oral users’ choice of technology.

Through the usability experiments we have established that oral users’ choice is affected by how the technology affords the users’ current situational norms at the time of interaction with the technology. Different cultures around the world have different cultural norms concerning loudness in speaking. Poyatos (1993) makes several examples of these various norms, for example, in Kenya and Ghana, shouting in the street and talking loudly indoors is unacceptable and Ghanaians find Nigerians too loud whilst loudness is a cultural characteristic in Spain and Italy. In addition to these cultural norms there are situational norms that fluctuate above or below the standard cultural norm as when people enter a large office where people

work quietly at their desks, in an exclusive lounge (particularly with an intimate low light level) or even during an abrupt silence whereby people immediately adjust their voice loudness accordingly. On the other hand, other environmental situations can force people to raise their voices, such as noisy factory or in a noisy party [14].

In BGR system the users chose the speech-enabled system because it affords the users the opportunity to shift easily from their standard cultural norm to the situational norm of loud soccer culture. Conversely, DTMF was chosen by its users because it enables them to shift easily to the situational norm of being mute about a stigmatized subject.

The loud culture of soccer evokes a situational norm that is above the cultural norm and this is attested the exuberant loud voice during participant interaction with the BGR system. On the other hand, the reality of stigma in HIV/AIDS illness within the socio-cultural norms of the OpenPhone's targeted users evokes a situational norm that is below the cultural norm as attested by the low, slow and monotonous voice during their interactions.

V. CONCLUSIONS

In both IVR systems the intended users perceived the systems as useful tools that would improve their livelihoods if the systems were to be fully deployed in the relevant regions. OpenPhone users viewed the technology as something that would improve their ability to deal with the illness at any time of the day and from their homesteads instead of having to travel to the clinic, which can be as far as 70 kilometres away for some users, in order to get help. The BGR users regarded their system as a helpful means that gives them results of recently played games whilst travelling or at work without having to wait for sports news on the television or the radio.

The DTMF modality's mute interaction affords the OpenPhone's intended users' requirement to interact with the technology in privacy and the modality also affords the users' shift from their standard cultural norm to the quiet situational norm that is demanded by the circumstances of their socio-cultural predicament of stigma. Similarly, the speech-enabled interaction modality affords the BGR users' need to interact with the technology in the loud manner that they are used to when interacting with other soccer fans. The speech-enabled modality also affords the BGR users to shift from the standard cultural norm to the loud situational norm that is congruent to the loud soccer culture in South Africa.

The researchers propose that the cultural conventions of the intended users need to be incorporated into the design of IVR systems for oral users as essential design attributes because these conventions are carried over into the interactions with speech technologies such as IVR systems, as attested by the reactions of the users when confronted with the different types of applications and modes of interaction.

ACKNOWLEDGMENT

The authors would like to acknowledge the assistance from the Rhodes University Centre of Excellence in Distributed Multimedia, Human Language Technologies Group at the Meraka Institute-CSIR, and the National Research Foundation.

REFERENCES

- [1] Barnard, E., Cloete, J.L., Coetzee, L., and Patel, H., (2003). Towards effective telephone-based delivery of government services. *9th International Winelands Conference*, Stellenbosch, 2003.
- [2] Barnard, E., Cloete, L., and Patel, H., (2003). Language and Technology Literacy Barriers to Accessing Government Services. In *Lecture Notes in Computer Science*, R. Traummüller, Ed., Springer-Verlag, Berlin, Germany, Issue 2739, pp. 37-42.
- [3] Brown, I., Cajee, Z., Davies, D., and Stroebel, S., (2003). Cell phone banking: Predictors of adoption in South Africa—An exploratory study, *International Journal of Information Management* **23** (5) (2003), pp. 381-394.
- [4] Chesney, M.A., Smith, A.W., Critical Delays in HIV Testing and Care: The Potential Role of Stigma, *American Behavioural Scientist*, Vol. 42, No. 7, 1162-1174 (1999).
- [5] Davitz, J. R., (1964). *The communication of emotional meaning*. New York: McGraw-Hill.
- [6] Duffy, L., (2005). Suffering, shame, and silence: The stigma of HIV/AIDS. *Journal of the Association of Nurses in AIDS care*, Volume 16, Issue 1, January 2005, pp. 13 - 20.
- [7] Gibson, J.J., (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin Co.
- [8] Hone, K. S. and Graham, R., (2001). Subjective assessment of speech-system interface usability. In *EUROSPEECH-2001*, pp. 2083-2086.
- [9] ISO 9241-11. Ergonomic Requirements for Office Work with Visual Display Terminals - Part 11. Guidance on Usability, 1998.
- [10] International Telecommunications Union, (2006). World Telecommunication Development Report 2006. *Mobile Cellular. Executive Summary*. Geneva, Switzerland.
- [11] Isomursu, M., (2008). In IFIP International Federation for Information Processing, Volume 272, Human-Computer Interaction Symposium, Peter Forbig, Fabio Paterno, Annelise Mark Pejtersen; (Boston: Springer), pp. 47-52.
- [12] Jokela, T., Iivari, N., Matero, J., and Karukka, M., (2003). The standard of user-centered design and the standard definition of usability: analyzing ISO 13407 against ISO 9241-11. In *Proceedings of the Latin American Conference on Human-Computer interaction*. Rio de Janeiro, Brazil, August 17 - 20, 2003. CLIHC '03, vol. 46. ACM, New York, NY, pp. 53-60.
- [13] Norman, D.A., (1999). Affordances, conventions, and design. *Interactions* Vol 6 Issue 3 (May 1999), pp.38 - 43. DOI=10.1145/301153.301168.
- [14] Poyatos, F., (1993). *Paralanguage: A Linguistic and Interdisciplinary Approach to Interactive Speech and Sound*. J. Benjamins, Amsterdam, Netherlands.
- [15] Reidpath, D. D. and Chan, K. Y. *AIDS Care*, Volume 17, Number 4, May 2005 , pp. 425-432.
- [16] SAFA: South African Football Association, (2010). About SAFA: Introduction to SAFA, [Online] Available: <http://www.safagoal.net/index.php?page=safaintro> [Accessed: 26 August 2010].
- [17] Scherer, K.R., (1979). Nonlinguistic vocal indicators of Emotions in Personality emotion and psychopathology. In and Psychopathology, C.E. Izard, (ed.), Plenum, New York, pp. 495-529.
- [18] Shackel, B., (1991). Usability—context, framework, definition, design and evaluation. In *Human Factors For informatics Usability*, B. Shackel and S. J. Richardson, Eds. Cambridge University Press, New York, NY, pp. 21-37.
- [19] Tractinsky, N., (1997). Aesthetics and apparent usability: empirically assessing cultural and methodological issues. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, United States, March 22 - 27, 1997). S. Pemberton, Ed. CHI '97. ACM, New York, NY, pp. 115-122.