Mobile User Experience for Voice Services: A Theoretical Framework

A. Botha; K. Caltreux; M. Herselman; A. Sharma and E. Barnard.

Abstract: The purpose of this paper is to provide a “Mobile User Experience Framework for Voice services.” The rapid spread of mobile cellular technology within Africa has made it a prime vehicle for accessing services and content. The challenge remains to provide these services and information on the technology that the user already owns and is proficient in using. To this end voice as service and distribution mechanism for information is, ahead of SMS and USSD, the most ubiquitous channel of access as it is available on all mobile phone handsets. User experience has been linked to the uptake and engagement with technology and services. As such it becomes imperative to acknowledge mobile user experiences for voice services in order to provide an optimal engagement opportunity that would facilitate participation by end users.

Keywords: Mobile User Experience, Voice services, Theoretical Framework

1 Introduction

The focus of this article is to provide a theoretical framework for mobile user experience (MEX) for the application of voice services. Before the framework can be presented it is imperative that we unpack a few concepts which are applied in this framework. Therefore an explanation of user experience (UX) will be followed by an explanation of mobile user experience with its specific components, as well as an explanation of what is meant by voice services in this context. The uniqueness of this framework is that it was adopted from a previously developed mobile UX framework and then applied to voice services specifically. This framework also forms part of an EU FP7 project (VOICES), initiated at the end of 2010.

User experience (UX), as a trans-discipline and emerging concept, has a multitude of definitions that are sympathetic to its origins, indicating its complexity and richness. Dix [1] observes that with the growth of the web, much software that traditionally sold as products have become services. He argues that where products allowed for one infrequent point of choice, services allow near continuous choice. As such, user experience becomes imperative to success. Pine and Gilmore [2] position experience as a unique offering of the emerging experience economy and argue that an experience occurs when an organisation intentionally uses services and goods to engage individuals in such a way as to create a memorable event.

With more consideration being paid to finding ways to reach the “bottom of the pyramid” [3] improved mobile access presents a unique opportunity for new voice services and voice service models. These are increasingly being facilitated by access to mobile cellular technology for individuals lacking large or predictable incomes [4]. User experience (UX) is an acknowledged factor in the successful uptake and use of technology [5-7], though it has received less attention for individuals in this category. Although user experience research informs practice on not only functional, but also hedonistic needs [8], limited empirical research is available to reflect this [9]. This is exaggerated when considering mobile user experience and even more so for voice services [10]. However, with the greater involvement of enterprises in developing countries in the provision of ICT related innovations [11] it becomes imperative to provide guidelines or frameworks that would, if not enable, increase the likelihood of uptake.

Practitioners and interaction designers cannot design, and much less control, a successful user experience. They can, however, design for one. As such, a framework for the mobile user experience for, in this case specifically voice services, becomes a 360° view of the mobile end user’s interaction with the mobile technology and the voice service.

2 Toward a Mobile User Experience

2.1 Focussing the User Experience

In order to understand user experience, Roto [8] argues that the term ‘experience’ encompasses many variables and that a focus on the interaction and experiences of the user with an interactive system is desirable. She states that: “[m]aking this distinction would help us to understand what is
meant by experience or UX, to identify the factors affecting user experience, and also to evaluate user experience in a systematic way [8]. She views user experience as a special case of experience that involves a service or product, relating to an interaction with the system, where the system does not need to be interactive. This paper will build on this understanding, and focus on a user experience as opposed to an experience.

Hassenzahl [12], referring to the end user mode of interaction with mobile technology, distinguishes between goal mode and action mode. Goal mode is characterised by the user wanting to achieve a goal. Action mode, on the other hand, is where the user is focused on entertainment. Entertainment activities include such interaction as browsing or gaming. The interactions that were considered for this paper are limited to goal-orientated interactions as opposed to general browsing or recreational interactions [12-15], mainly because a user will interact with a voice service on a mobile device to reach a specific goal.

2.2 The User Experience
There is little consensus in literature on either the definition or characteristics of a ‘user experience’. Literature does, however, generally agree that a UX would include subjective attributes and social aspects. These subjective attributes and social aspects would be additional considerations in a space that has previously concerned itself mainly with ease-of-use and implies considerations that are beyond the task-related [1, 9, 16, 17]. Preece et al. shape the HCI concern and state that: “The dominant framework that has characterized HCI has been cognitive. In general, cognition refers to the processes by which we become acquainted with things or, in other words, how we gain knowledge. These include understanding, remembering, reasoning, attending, being aware, acquiring skills and creating new ideas” [18]. Hassenzal and Tractinsky reiterate this stating: “Since its early days, HCI research focused almost exclusively on the achievement of behavioural goals in work settings. The task became the pivotal point of user-centred analysis and evaluation techniques (e.g. usability testing). To ensure the interactive product’s instrumental value became the major endeavour of the field” [9].

There are several reasons for the illusiveness of a universal definition of UX. The first can be ascribed to the broad range of vague and dynamic concepts on which there is little consensus regarding the inclusion or exclusion of attributes. The second reason concerns the unit of analysis for UX, which ranges from a single aspect of an individual user with a standalone application, to all aspects of multiple users with many and diverse services and applications across domains. The third has to do with the fragmented research focus [19]. The various definitions articulated in literature [8, 20-27] all directly or indirectly reflect the findings of the review done by Hassenzahl and Tractinsky [9]. They identify three high-level components that affect the user experience, namely the user, the system, and the context.

Although the identified elements of the UX remain for mobile interactions, there are added complications and dimensions due to the mobility of the interaction and the personal nature of the technology [25]. These are reflected on in the following section.

2.3 The Mobile User Experience (MEX)
The literature that reflects on the mobile user experience is mostly limited to expert opinions and insights gained from solution-driven interventions. Ledfort [28] argues that the mobile user experience is not monolithic as it is dependent on a number of factors. Elements of context, networks, and the business-related issues such as cost, and unique affordances of mobile users and use have been suggested as additional considerations [15, 29-35]. Planning for a mobile user experience would imply the optimal consideration of additional components that impact on the MEX from the MHCI considerations and from the voice interaction considerations. The components that would frame the MHCI are identified as mobile users, mobile devices, mobile networks, mobile business processes and mobile use.

An overview of each of these focus areas is beyond the scope of this paper and limits the outline to conclude that there are many challenges and potential solutions for effective interaction with mobile devices and services. However, these solutions are underpinned by common components that make up the interaction as outlined in the structure for MHCI.

The components of a user experience, outlined in section 2.2 as the user, the system and the context can now be expanded on to include additional considerations towards a Mobile User Experience as:
User: The mobile user; Mobile use.

System: Mobile device; Mobile business practices; Network affordances; Mobile Applications; Mobile Interaction

Context: Mobile Context

From these considerations a comprehensive Framework was presented by Botha to outline the factors or each of the components and their impact. These were adapted to incorporate specifics related to voice services. The next section investigates voice services focussing on goal driven interactions towards presenting a framework for mobile user experience for voice services specifically.

3 Voice Services

In recent years there has been a significant body of work generated in the voice-based services area such as [37-42] to name a few, which covers various domains such as education, health, agriculture, finance, etc. Voice-based services are often referred to as spoken dialogue systems (SDS) or interactive voice response (IVR) systems in literature. Both allow a user to access information or a service via the voice channel of their mobile phone, by navigating through voice menus where input by the user is through speech (for SDS) or dual tone multi frequency (DTMF) (for IVR). The user typically interacts with the service through the means of a simple telephone call to the service’s phone number or more recently as proposed by Google [43] through a voice-based search using the data channel of their mobile phone.

Numerous communities in developing world regions face barriers to information/service access, including infrastructure, distance, language and literacy. Many government entities and non-profit institutions need to deliver services and provide timely, accurate and relevant information to their communities of interest, which can be a challenging task due to these barriers. Voice-based services can play an important role in addressing these barriers and bridging the information gap as mobile phones are by far the most widespread form of ICTs in developing world regions [11], [10]. Being independent of mobile phone device type and operator is an added advantage of voice-based services [10]. Barriers of language and literacy are also addressed as the service’s content can easily be made available in local languages and most users are comfortable with the concept of making a telephone call as opposed to operating a PC or interacting with a mobile interface. Voice-based services also have the further advantage that they do not require any computer infrastructure from the user end and can be used from anywhere, alleviating transport-related costs and delays.

Plauche et al. [44] developed one of the first voice-based services for low literacy users in the agriculture domain. It was found that low literacy users were able to navigate a SDS but with differences in task completion for low literate and illiterate users. The topic of input modality in voice-based services has also been explored by Grover et al. [37], Sherwani et al. [38], Patel et al. [39, 45], and Lerer et al. [46]. Results have varied in terms of user performance (task success) and user preference for any particular modality. Sherwani et al. and Lerer et al. [46], found speech input provided a significantly higher task success rate than DTMF. Conversely Patel et al. [45] found that user performance was better with DTMF input, whilst Grover et al. [37], report no significant difference in user performance between speech and DTMF input. For user preference, both Grover et al. [37] and Patel et al. [45] report that users preferred DTMF over speech input, whilst Sherwani et al. [38] report no significant difference in user preference, and Lerer et al. [46] although not explicitly reporting user preference suggest that users did not like the DTMF aspects of the system. Its noteworthy to mention that though all these studies targeted developing world users, they were conducted in different domains and contexts with different types of users; HIV info (Grover et al.), agriculture info (Patel et al.), general health info (Sherwani et al.) and an audio survey (Lerer et al.).

A language learning service by the BBC, termed “BBC Janala” [47] provides English language lessons via an IVR in Bangladesh. The user “dial[s] up a series of three-minute-long English lessons for 3 taka (2.5 pence) [USD 0.04] each, which is less than the cost of a cup of tea at a roadside stall in Dhaka” [47]. The service received over 750,000 calls in the first month [47] and to our knowledge is one of the rare successful examples of a ‘user-paid’ voice-based service in the developing world. Another notable example of a ‘user-paid’ voice-based service is that of “Lifelines India” [42] which provides a question-answering service in the agriculture and education domains. Through the assistance of a community field worker, the user calls the IVR, records their query and obtains a unique query ID number. On the back-end, within 24 hours a knowledge worker posts the
response to the query using expert opinion and a knowledge database of similar FAQs. Later, the user retrieves the query’s response by dialling into the IVR again, with the cost of the call being 5 rupees (USD 0.11).

Patnaik et al. [48], Medhi et al. [49], and Kote et al. [50] compare a range of mobile user interfaces (UI) which include voice-based services. In [48], it was found that error rates for data collection through a live operator (voice) were significantly lower as compared to SMS, and electronic forms (via mobile phone). Medhi et al. [49], compare text-based interfaces such as electronic forms, SMS and USSD with text-free interfaces such as an SDS, graphical UI and a live operator. They report that “textual interfaces were unusable by low literacy users and difficult to use by novice users”. In the case of text-free interfaces, the live operator (voice) was found to be the most effective, with varying results for voice and graphical UIs respectively. In particular for voice UIs, Medhi et al. suggest that users, who are somewhat familiar with the concept and the general terminology (prompts) of the voice UI, were faster and more independent in their task execution. However, overall, graphical UIs had a higher task completion rate but users took significantly more time to complete the task and required more prompting and encouragement during the study. In a similar vein, Kote et al. found that users preferred an IVR over SMS for a service that crowd sources water availability information India.

Agarwal et al. [41] discuss the adoption of four different pilot deployments of voice-based services for general community information and agriculture in India. The authors highlight that, choosing a local partner organisation within the user community to act as an intermediary and ensuring content is relevant and moderated where required, are important factors that play a role in the adoption of voice-based services. Similar findings on content and adoption were reported by Grover et al. [51] for “Lwazi” which piloted a voice-based community information service for managers of government community centres, in six areas across South Africa. In [51] it is reported that the availability of information sources which provide content for the service was a crucial factor, and in terms of multilingualism in developing world environments, a dominant language or two usually prevails in an area with most users tending to be multilingual and conversant in the dominant language(s).

Across the various studies described above, numerous significant findings have been reported and several recommendations on design and deployment have been made. However, to date an overarching framework that explicates the space of user experience for voice-based services in the context of a mobile phone user has not emerged. Some noteworthy studies that provide an overview of a number of salient factors to be considered and recommendations for the development of voice-based services include those of Barnard et al. [52] and Grover et al. [40, 53]. In this paper we further expound on these recommendations and those of the numerous voice-based based services mentioned above and present a consolidated framework on mobile user experience for voice-based services.

4 Framework for Mobile User Experience in Voice Services

This framework, which was adapted from Botha [36-38], consists of components of mobile UX or MEX which were outlined in section 2.3 as well as MEX factors and how these impact on mobile and voice contexts of use. Colour is used to reflect the following contexts:

<table>
<thead>
<tr>
<th>Component</th>
<th>Mobile user experience factor</th>
<th>Impact in mobile and voice contexts of use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile User</td>
<td>Mobile Users have unique characteristics [8, 15, 35]</td>
<td>The user occupies multiple social spaces simultaneously. [33, 54-60]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The user is distracted (short attention span) [34, 35, 55, 56]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The user multitasks [35, 54, 55, 61]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The user is available or considered as connected [55, 56]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The user is contextual and the environment affects device use. [33, 34, 54-56, 58, 59]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The user personalises the device [34, 55, 56, 61]</td>
</tr>
</tbody>
</table>

Evidence of relevant literature for each impact factor is then provided to develop the theoretical framework and a discussion or interpretation of the influences of mobile, voice and both mobile and voice is then provided.
<table>
<thead>
<tr>
<th>Component</th>
<th>Mobile user experience factor</th>
<th>Impact in mobile and voice contexts of use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile Use</td>
<td>User appropriation of the technology-in-use is facilitated [61]</td>
<td>The technology is convenient to use (available) [61]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The user is in control of the mobile device (shared and multiple usage) and the speech application [52, 61, 62]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The user considers the device fashionable [25, 53, 61, 62]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The user considers the technology fashionable/as a status symbol [39]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The user can identify with the technology-in-use as “our stuff” [61]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The user is exposed to long-term engagement with the application [40, 51]</td>
</tr>
<tr>
<td>Hedonic experience of use is facilitated [9, 32-34, 64]</td>
<td>The user enjoys using the mobile device, [9, 55, 56]</td>
<td>The user will use the mobile device again, [9, 22]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The user does not experience frustration [64]</td>
</tr>
<tr>
<td>Mobile Device</td>
<td>Device capabilities support the interaction adequately performance issues (Hardware) [10, 65-67]</td>
<td>The display is clear and visible and accessible during the interaction (e.g. sunlight) [66-68]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Display is capable of rendering content for interaction [69]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Battery life is adequate to support the required mobility [66-68]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The memory capacity is sufficient. There is sufficient capability to extend the memory if needed [66-68]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The device processing power supports the interaction sufficiently [66-68]</td>
</tr>
<tr>
<td></td>
<td>The imbedded software support the interaction adequately (Software) [10, 55, 68]</td>
<td>Functionalities adequately enable the interaction [68, 70]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Functionalities of the device are usable for the interaction. [56, 68]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Operating system supports installation of application used in the interaction [56, 68, 71]</td>
</tr>
<tr>
<td></td>
<td>Software is usable in use [14, 64, 72]</td>
<td>Functionality feedback is understood [56]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The software embedded in the device is error free [64, 72]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The interaction with the software embedded in the device is easy to remember [62, 64, 72]</td>
</tr>
<tr>
<td></td>
<td>Mobile Business Practices [10, 15, 35, 55, 56, 74-76]</td>
<td>The software embedded in use is easy to learn [62, 64, 73]</td>
</tr>
<tr>
<td></td>
<td>Network is available [10, 15, 35, 55, 56, 78-82]</td>
<td>There is network coverage [36, 62]</td>
</tr>
<tr>
<td></td>
<td>The network does not need network coverage [34, 55, 56]</td>
<td>The interaction is not needed [35, 55, 75]</td>
</tr>
<tr>
<td></td>
<td>Network is reliable [10, 55, 56]</td>
<td>Can perform the expected service dependably, accurately and consistently [35, 55, 83]</td>
</tr>
<tr>
<td></td>
<td>Network facilitates interaction [35, 75]</td>
<td>Network services are sufficient to support interaction [35, 55, 56, 75]</td>
</tr>
<tr>
<td></td>
<td>Mobile Interaction supported by usability of application. [15, 55, 56, 64, 84]</td>
<td>(usability of application) [55, 56]</td>
</tr>
<tr>
<td></td>
<td>The Mobile Application supports the interaction [10, 34, 55, 56, 86, 87]</td>
<td>Service or product is simple and easy to use (ease of use) [15, 62]</td>
</tr>
<tr>
<td></td>
<td>Mobile Application accesses the interactions that are native to the phone [10, 34]</td>
<td>Important functionalities are easy to find (fluency of navigation) [15]</td>
</tr>
<tr>
<td></td>
<td>Provides service and content to user when needed [15, 34, 55, 56]</td>
<td>Interaction needed in application is learnable [62, 64]</td>
</tr>
<tr>
<td></td>
<td>Provides service and content to user where needed [15, 34, 40, 55, 56]</td>
<td>Interaction is safe and secure [64, 84]</td>
</tr>
<tr>
<td></td>
<td>Mobile Application makes task easier [15, 34, 40, 62]</td>
<td>Interactions are suited to mobility e.g. One hand information input on the move [77, 85]</td>
</tr>
<tr>
<td></td>
<td>Mobile Application provides only useful information during interaction [15, 56]</td>
<td>Impact in mobile and voice contexts of use</td>
</tr>
<tr>
<td></td>
<td>Mobile Application provides appropriate functions for interaction [34, 40, 56, 62, 64]</td>
<td>Application is reliable and performs service dependable, accurately and consistently [34, 55, 56]</td>
</tr>
</tbody>
</table>
The BBC Janala service provides degrees of involvement of local users developing world contexts is the provision of content in the local language (and accent), as well as trustworthy, relevant, reliable, timely, and trustworthy for a positive user experience to occur. A critical success factor in developing world contexts is the provision of content in the local language (and accent), as well as involvement of local users in the design of the application [62].
As mentioned in section 3, developments in deploying voice-based services such as voice search on mobile devices are resulting in the data channel on the mobile device becoming an impact factor in voice services. Functionalities and operating systems on the mobile device which enable interaction with the application and are usable for the interaction, are therefore becoming critical to the success of such services and the user’s experience of such services. Barnard et al. [43] discuss voice-search systems for development in more detail.

Mobile network infrastructure is also a critical impact factor for mobile voice services UX, with the emphasis on a reliable network with sufficient network services to support the interaction.

Finally, particularly in developing world contexts, the costs relating to voice services and the implications for mobile business practices have an impact on UX of mobile voice services, with value for money being critical. The charges (five Rupees) described in section 3 for the LifeLines India Agriculture service, serve a dual purpose: the users value and respect the information because it is not totally free; and it enables the creation of a sustainable business model. Lall [42] reports caller satisfaction levels of 96%, profit increases of between 25% and 150% and a rise in call volumes from “100 calls per month at launch, to an average of 350 calls daily”. A sustainable business model further relies on stakeholders understanding their roles in the deployment of the service, and the service being aligned with or supplementing existing information or services channels [42].

Acknowledging the above factors as being critical to mobile user experience for voice-based services, enables their use as a starting point for developing a set of metrics to measure user experience of these services.

6 Conclusions
In this paper we give an overview of user experience in using mobile technology to interact with voice-based services. We indicate that a theoretical framework developed for mobile user experience can be adapted and applied to voice-based services. This adaptation indicates that in most instances, there is an overlap in factors that impact on user experience of mobile technologies and of speech technologies. The framework is expanded when impact factors pertaining specifically to speech-driven services are added.

Our future research will involve applying this theoretical framework to real-world applications. First, using the critical factors described above, we will distil from the theoretical framework, metrics for measuring mobile user experience for voice services. These metrics will then be used to measure user experience in two speech applications as case studies. We will report on the findings of this research in future papers.

7 Acknowledgements
This research is partially funded by the European Union through the 7th Framework Programme (FP7) under grant agreement Num. 269954.

The authors would further like to acknowledged CSIR Meraka and Department of Science and Technology.

8 References
29. Bevan, N. Extending quality in use to provide a framework for usability measurement. in HCI International. 2009. San Diego, California, USA.


