Mobile Search Interfaces for isiXhosa Speakers: A Comparison Between Voice and Text

Morebodi Modise
Department of Computer Science
University of Cape Town
Cape Town, South Africa
modisemorebodi@gmail.com

Hussein Suleman
Department of Computer Science
University of Cape Town
Cape Town, South Africa
hussein@cs.uct.ac.za

***Abstract*—Search interfaces have for a long time been targeted at the resource-rich languages such as English. However, due to the increase in use of mobile phones in developing countries, these interfaces can now be adapted to languages in these settings to support information access on the Web. In this study, we propose two mobile search interfaces - text and voice - to support isiXhosa speakers to search for information on the Web. Experiments were conducted with 34 native isiXhosa speakers to measure satisfaction with the two interfaces. The results show that isiXhosa speakers were more satisfied with the mobile text interface.**

***Keywords*—mobile search interfaces; developing countries; mobile phones; isiXhosa**

I. Introduction

The use of mobile phones in developing countries has increased significantly in recent years. As of 2014, there were 6.9 billion mobile cellular subscriptions, of which 78% were from developing countries [1]. This has made it possible for people in these settings to have an alternative means to the desktop computer that enables them to search for information on the Web. Mobile phones can be extended to better address information needs for users in developing countries who search for information on the Web using these devices [2]. However, searching for information on the Web using mobile phones presents several challenges and limitations such as smaller display, which limits the amount of information to be displayed at a time, and insufficient battery life needed to support the search process [3]. These challenges and limitations present an opportunity for further research, especially given that these devices also offer additional benefits compared to their desktop counterparts such as searching the Web while mobile.

While interacting with information on the Web using mobile phones, information seekers make use of search user interfaces commonly referred to as search interfaces, a notion that will be adopted throughout this study. Most search interfaces today focus on supporting English literate users [4]. As such, they have become unusable to some communities in developing countries where English literacy is considerably low, in spite of accessibility to mobile phones [5]. For example, in most African contexts, users may own a mobile phone and may not be fluent in English but rather in their mother tongue only. Because isiXhosa speakers reside in a developing country, where it might not always be possible to have access to a PC, use of mobile phones is proposed as an alternative.

This study focuses primarily on comparing two mobile search interfaces: text and voice - to support isiXhosa speakers to search for information on the Web using isiXhosa as the discovery language.

II. Related Work

A. Language Support

1) Mobile Text Interfaces: In an effort to support access to non-English languages' informational sources that are rapidly increasing on the Web, mobile text interfaces have been adapted to other languages, including low resourced languages [4]. Hattab et al. [6] present an Arabic search engine called Addaall, which uses different levels of Arabic morphological knowledge to enhance the quality of the search engine. The results of the study have shown promise that the study's approach is practical and it offers a significant improvement to the search engine, therefore its interface.

Ricardo et al. [7] describe the need to comprehend what people around the world search for in their mobile devices so as to design the best interface to support them. Further, they investigated the usage of mobile devices in searching for information in Japanese by collecting sample log data from Yahoo!. The collected data was from both the mobile and desktop query logs, as provided by Yahoo! in 2006. The results of the study show that mobile queries are shorter than desktop queries, which is mainly due to the text-input restrictions caused by the small display on mobile devices.

2) Mobile Voice Interfaces: Mobile voice interfaces have also been extended beyond the resource-rich languages like English to other languages in both the developed and developing countries. Schuster and Nakajima [8] highlight problems and solutions they encountered while building the mobile voice search system for the Japanese and Korean languages. One of the challenges they faced was adapting words, numbers and symbols from the English language to the Japanese and Korean languages, as these languages have complicated character sets. The results of the study have shown that they

were successful in building a mobile voice search system for the Japanese and Korean languages at Google. Based on their results, they proposed a simplified standard for building mobile voice search systems in other international languages. Similarly, Shan et al. [9] carried out a detailed investigation into searching on a mobile phone using voice in Mandarin Chinese. They detail the process of collecting audio data for building language, pronunciation and acoustic models. The results of the study reflect a successful voice search mobile application, which has been well-received by end users and performs well with different user accents.

Sung et al. [10] detail an attempt to build a mobile voice search system for the Cantonese language, a language spoken in the southern parts of China. They found out that most users find it ideal to speak rather than type out search queries when searching for information on the Web. This is because the Cantonese characters are difficult to type out on a mobile phones' smaller keyboard. In associating voice search with the potential of improving lives in the developing world, Barnard et al. [11] investigated the impact of speech technology, in particular voice-search, in improving the presence of speech technology in the developing world. They argue that speech technology is necessary to help address some of the challenges faced in the developing world, such as illiteracy and the ability to access information. The results of the study highlight a successful journey in building a mobile voice search system in two of South Africa's eleven official languages, namely: Zulu and Afrikaans.

III. System Design and Implementation

We followed an iterative 4-step design approach to design and develop the two mobile interfaces used in this study, as shown in Fig. 1 [12]. The first step involved designing and, developing each interface according to the design guidelines suggested during a design focus group by 4 native isiXhosa speakers, and best design practices as recommended in the literature [13][14][15][16]. The next step involved the developer testing the usability of each interface, thereafter passing it on to fellow researchers in the ICT4D laboratory to do the same. As the testing phase was underway, design and feature suggestions were made and, where possible, they were incorporated into the design.

The refinement stage involved revisiting the design of the interfaces based on design modifications suggested during the testing phase. This resulted in modifying key features to make them as close as possible to the general requirements of search interfaces [16]. In the evaluation stage, target users played a vital role in ensuring that the interfaces were well suited to isiXhosa speakers. At the end of the evaluation, the design process was repeated, including new features.

A. Prototype

The end result of the design was a mobile prototype, which had two mobile interfaces developed using the Android platform, targeting Android 3.0 (Honeycomb) and above. The main communication language in each interface is isiXhosa, as this is the target group's language. For experimental purposes, the two interfaces were separated into two mobile applications to make a clear distinction between the two interfaces to participants in the study. Additionally, both interfaces needed an Internet connection in order to fetch the search results from a search engine server hosted in the Department of Computer Science at the University of Cape Town, South Africa. In the case where Wi-Fi was not available, mobile data was used.

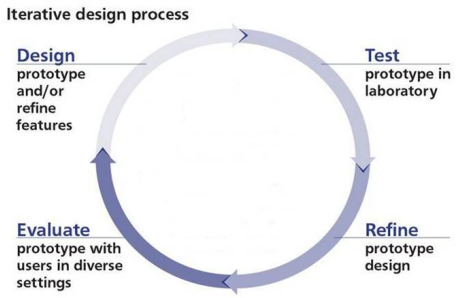


Fig. 1: Iterative design used to develop the two mobile interfaces

The mobile text interface allows the user to search for information by typing in the search term in the white search box shown in Fig. 2a, thereafter pressing the search button to retrieve a list of search results in text, as shown in Fig. 2b. The user can then scroll through the results and access those most relevant to their search, as shown in Fig. 2c. In the mobile voice interface shown in Fig. 3a, the user taps the microphone icon and speaks a search term, where the interface attempts to detect every spoken term, as shown in Fig. 3b. Thereafter they obtain a list of the most relevant search results in voice. The user can opt to listen, replay, mark results as relevant or skip through results as they are being played, using the buttons shown in Fig. 3c. The two orange arrow buttons provide back and forth navigation through the list of voice results. The tick button marks the currently listened to result as relevant.

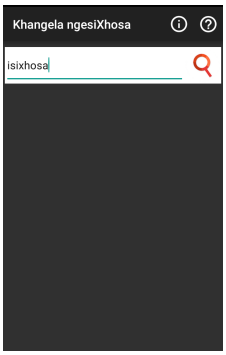
In regard to the voice interface, the objective was to use a state-of-the-art speech recognition system, which could handle noise interference, a topic that is out of scope for this study.

IV. Experiment Design and Evaluation Methods

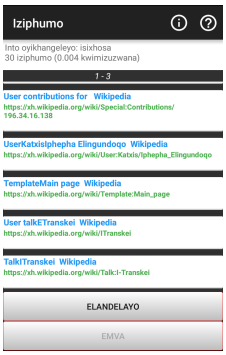
A. Participants and Sampling

Participation was on a voluntary basis, which meant participants participated in the study by choice and were free to withdraw participation at any given time. The design focus group of 4 participants was selected to come up with the design of the two mobile interfaces: text and voice. The objective was to have the interfaces co-designed with native isiXhosa speakers. This sentiment is supported by the literature that emphasizes the need to engage target users in the design process of interfaces, starting from the initial design until the end product is achieved [17][15].

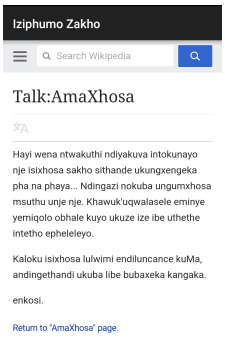
This study adopted a convenience sampling strategy because most of the participants who met the requirements of the study were already accessible within the University. The recruitment



(a) Query submission interface

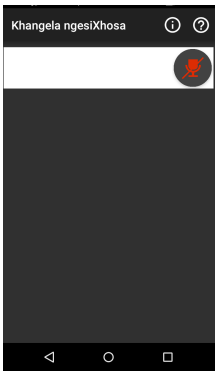


(b) Results interface

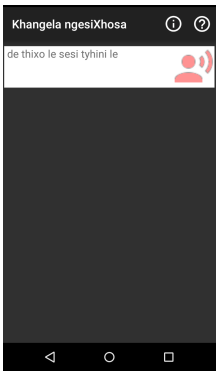


(c) Single result view

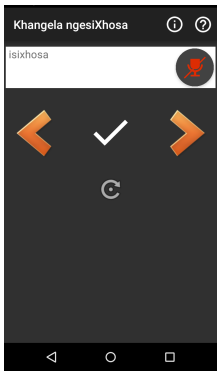
Fig. 2: Mobile text interface



(a) Query submission interface



(b) Detecting voice queries



(c) Detected voice query with list of voice results

Fig. 3: Mobile voice interface

process was initiated by sending out an email to the entire University, targeting native isiXhosa speakers to invite them to participate in the experiment. A total of 3 native isiXhosa speaking students were recruited for the pilot study and a total of 34 native isiXhosa speaking students were recruited for the main experiments. The goal was to test the feasibility of the interfaces using participants who spoke the isiXhosa language and were familiar with the search concept rather than testing search with a population where some people were not used to the search concept. The number of participants needed for the main experiments was computed using SurveyMonkey's[1] sample calculator. Additionally, to ensure a large turnout of participants: a) participants were invited to the study as close to the experiment date as possible, and b) each participant was given a R40 cash incentive for participating in the study.

B. Data Collection

Electronic questionnaires were presented to the participants to get their feedback regarding the two interfaces. These were electronic questionnaires completed using a laptop, and were constructed and provided for use in the study by the Software Usability Measurement Inventory (SUMI). SUMI[2] is a tried-and-tested method of obtaining information about software quality as viewed by the user [18]. SUMI is scored and interpreted with reference to a standardization database of about 2000 profiles ranging from office systems and communications software to Computer-Aided Design packages [19].

The SUMI questionnaire contains 50 questions, from which participants had three options to select, namely: agree, disagree and undecided. The agree option implies that the participants feel the questionnaire statement relates to their experiences with the interface while the disagree option means the user feels the questionnaire statement does not relate to their experiences with the interface. The undecided option implies that the participants could not decide on the answer to the given questionnaire statement. SUMI provides analyzed results in two forms: firstly, the Global scale score reflects the general usability factor of the given interface; and, secondly, there are 5 usability subscales: a) Helpfulness b) Affect c) Control d) Learnability and e) Efficiency [19].

Helpfulness measures whether the given software is self-

[1]https://www.surveymonkey.com/mp/sample-size-calculator/

[2]http://sumi.cc.ie/whatis.html

explanatory and has enough help facilities such as documentation and help menus. Affect measures the user's emotional behavior towards a given software, that is, the likability aspect. Control measures the degree to which the user feels in command of the interface as opposed to the software controlling the user while performing a given task. Learnability measures the adaptiveness and speed with which the user feels they were able to master the given software. Efficiency measures the degree to which the user feels the given software helped them to do their tasks.

The Global scale is measured by 25 questionnaire statements from the 5 subscales. The Global scale represents the single construct of perceived quality of use rather than a mere average of all the 50 questions on the questionnaire [18]. An average software system is set by SUMI to a score of between 40 and 60 on the Global scale and all the 5 subscales discussed above [18]. Any software that rates less than 40 is considered below average and any software that rates above 60 is considered above average.

A single SUMI questionnaire was used on each interface. In order to separate questionnaires by their respective interfaces, each questionnaire had a unique password corresponding to a particular interface. The passwords were entered by the researcher prior to each participant answering the questionnaire. Participants took an average of 10 minutes to answer the 50 questions on a single questionnaire. Therefore, it took each participant about 30 minutes to complete answering questionnaires for the two interfaces.

C. Conduct of the Experiment

1) Environment: All parts of the experiment were conducted one after the other in a quiet and comfortable guest lecture room in the Computer Science Building at the University of Cape Town. Although the mobile text interface did not require a quiet environment, it was also evaluated in the same room to ensure consistency in the experiment.

2) Search Tasks: A search task per participant was made up of 10 search queries, as shown in Table I. The queries were printed on an A4 page without their related English translations, and a brief overview of the project was highlighted in the Informed Consent forms. The participants were notified that there are many relevant documents per query, and as such must try to find as many of these documents as they could.

3) Pilot Study and Experiments: A pilot study was conducted prior to the main experiments. The pilot study was conducted to verify the practicality of the experimental protocol for this study.

The within-groups experiment approach was adopted during experiments, where each participant evaluated each of the interfaces using the search tasks discussed above. The first participant first evaluated the mobile text interface, and at the end of the evaluation completed a corresponding SUMI questionnaire. Then, they evaluated the mobile voice interface, and at the end of the evaluation they completed the corresponding questionnaire. The second participant evaluated the mobile voice interface then finally the mobile text interface.

TABLE I: isiXhosa search queries with their respective English translations

isiXhosa Query	**English Translation**
isiXhosa	The isiXhosa language
emzantsi afrika	South Africa
inkxaso	Support of the females / Support from the females
ndihlala emthatha nabantwana	I live in Mthatha with my children
isebe lezemihlaba eMpumalanga	The Department of Land Affairs in Mpumalanga
amanzi	Water
isixeko sasekapa	City of Cape Town
umhlaba	Soil
hamba uye emaXhoseni	Go to the land of the Xhosa people
ndikhathazekile ndoda	I am a troubled man

These alternations were applied until the last participant, that is, alternating between the mobile text and mobile voice interfaces. In each part of the experiment, each participant was given the same search tasks and completed the SUMI questionnaire at the end of the evaluation.

The following guidelines were applied during experiments:

- Each participant was given an overview of the study as well as the purpose of the experiment.
- Any questions regarding the research or the experiment were answered to the participant's satisfaction.
- An informed consent form was given to the participant, and the participant guided on how to complete it.
- The participant was given printed search tasks to evaluate the interfaces.
- The participant was then handed a mobile phone with the text interface to evaluate using the search tasks. Upon completion, the participant answered a SUMI questionnaire for the mobile text interface.
- The participant was given another mobile phone with the mobile voice interface to evaluate using the same search tasks used to evaluate the mobile text interface. Upon completion, the participant answered a SUMI questionnaire for the mobile voice interface.
- Finally, the facilitator thanked the participant as a sign of appreciation, and handed them the R40 cash incentive.

D. Evaluation Criteria

To address user satisfaction, SUMI questionnaires were used to collect qualitative feedback, as provided by the participants in the study. This was used to measure satisfaction based on the SUMI Global scale as well as the 5 subscales, namely: Learnability, Control, Efficiency, Helpfulness and Affect.

A paired t-test was used to compare the mobile interfaces. This is a statistical test used to compare two population means where one has two samples in which observations in one sample can be paired with observations in the other sample [20]. A significant difference in the compared interfaces would imply the interface with a higher SUMI Global scale rating is the most satisfying.

V. Results and Discussion

34 participants participated in the experiments. However, 33 responses were analyzed for the mobile voice interface and 32 for the mobile text interface as some participants failed to submit questionnaires due to Internet failure during submission time, as the questionnaire was administered online. The consolidated SUMI results for the two interfaces are presented in Table II.

TABLE II: Statistical comparison of the mobile text interface against the mobile voice interface on the SUMI scales

	Mobile Text Interface		Mobile Voice Interface		Statistical Comparison	
SUMI Scales	Mean	St Dev	Mean	St Dev	t-test Value	p-Value
Global	58.19	8.14	49.7	12.07	3.314	0.001
Efficiency	56.91	12.24	45.39	14.38	3.473	<0.001
Affect	58.75	11.21	50.06	12.52	2.945	0.002
Helpfulness	54.22	8.31	52.12	9.88	0.926	0.179
Controllability	53.72	10.0	49.58	8.53	1.798	0.039
Learnability	57.91	9.21	51.27	11.87	2.514	0.007

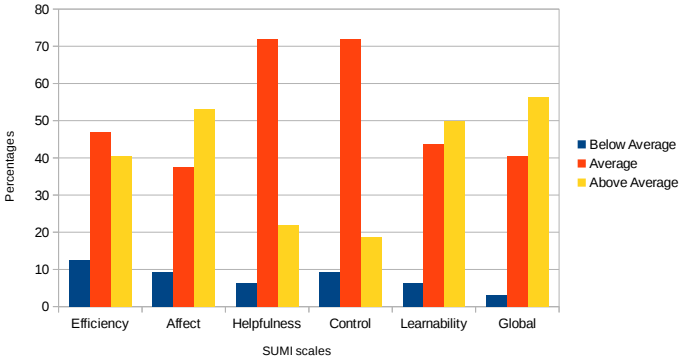


Fig. 4: Distribution of responses on the Global scale and 5 SUMI subscales for the mobile text interface

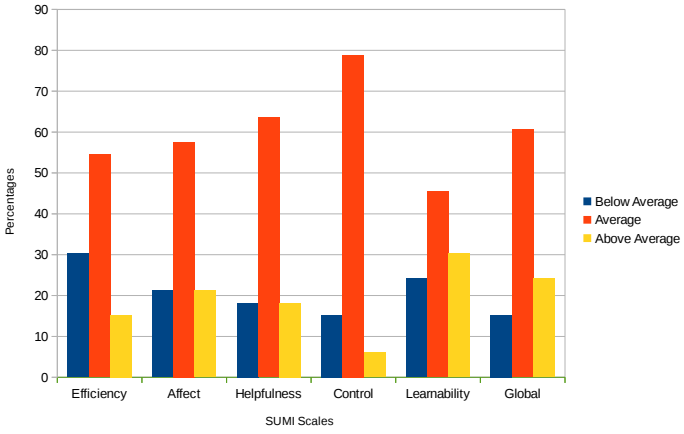


Fig. 5: Distribution of responses on the Global scale and 5 SUMI subscales for the mobile voice interface

1) Efficiency: According to the results in Table II, there was a significant difference between the Efficiency of the mobile text and voice interfaces. This implies that the text interface was more Efficient than the mobile voice interface. The text interface had an Efficiency mean of 56.91 and the voice interface had 45.39, as shown in Table II.

Further evidence in Fig. 4 shows that 46.9% of the participants rated the text interface's Efficiency as average while 40.6% rated it as above average. Even so, the voice interface's Efficiency was rated as average by 54.5% of the participants while 15.2% rated it as above average.

These statistics show that most of the participants were happy with the extent to which the interfaces assisted them in the search process, although the mobile text interface was presumed more efficient.

2) Affect: From Table II, there was a significant difference between the Affect of the mobile text and voice interfaces. This shows that the text interface was liked more over the voice interface. Additionally, Fig. 4 shows that 53.1% of the participants rated the text interface's Affect as above average, 37.5% rated it as average and 9.4% rated it as below average. However, 57.6% of the participants rated the voice interface's Affect as average and 21.2% rated it as above average and below average.

These figures also show that the participants liked both interfaces despite preferring the text interface, with the text interface having a mean of 58.75 and the voice interface having 50.06.

3) Helpfulness: The results in Table II show that there was no significant difference in the Helpfulness of the text and voice interfaces. The text interface had a Helpfulness mean of 54.22 while the voice interface had 52.12. This implies that both the text and voice interfaces' Helpfulness was considered equivalent by participants of the study. Yet, Fig. 4 shows that 71.9% of the participants rated the text interface's Helpfulness as average while 21.9% rated it as above average and 6.3% rated it as below average. In regard to the voice interface, Fig. 5 shows that 63.6% of the participants rated the voice interface's Helpfulness as average while 18.2% rated it as above average and below average.

It is obvious from the above statistics that most of the participants in the study found the interfaces to be self-explanatory with enough help facilities. The below average percentages of 18.2% for the voice interface and 6.3% for the text interface, however, suggest that improvements are still required to ensure the interfaces better help isiXhosa speakers.

4) Controllability: According to the results in Table II, there was a significant difference between the Control of the text and voice interfaces. This demonstrates that participants found it easier to command the text interface rather than the voice interface. As shown in Fig. 4, 71.9% of the participants rated the text interface's ability to be controlled as average, 18.8% rated it as above average and 9.4% rated it as below average. On the contrary, 78.8% of the participants rated the voice interface's ability to be controlled as average while 6.1% rated it as above average, as seen in Fig. 5.

5) Learnability: Similarly, there was a significant difference between the Learnability of the text and voice interfaces.

The text interface had a Learnability mean of 57.91 while the voice interface had a mean of 51.27. From Fig. 4, 50% of the participants rated the text interface's Learnability as above average, 43.8% rated it as average and 6.3% rated it as below average. However, as seen from Fig. 5, 45.5% of the participants rated the voice interface's Learnability as average while 30.3% rated it as above average and 24.2% rated it as below average.

These statistics demonstrate that native isiXhosa speakers were able to adapt better to the functionality of the text interface compared to the voice interface.

6) Global: The Global scale rating provided the general usability satisfaction scores of the two interfaces[3]. It is evident from Table II that the text interface's usability was more satisfying than that of the voice interface. In addition to the results of the subscales discussed above, the text interface was rated by 56.3% of the participants to be above average while 40.6% felt the interface was average, as shown in Fig. 4. Only few participants (3.1%) felt the interface's usability was below average. Nonetheless, the results from Fig. 5 shows that 60.6% of the participants rated the voice interface's usability as average, 24.2% rated it as above average and 15.2% rated it as below average.

From the results above, it can be seen that the mobile text interface was easy to learn, supported the participants with appropriate speed, and was easy to command.

VI. Conclusion

This paper reports on an experiment that set out to provide native isiXhosa speakers with appropriate mobile search interfaces. The goal was to compare text and voice interfaces to determine which is preferable, noting that recent research and mobile phone innovations suggest that voice interfaces could be the preferred choice.

A user-centred design process was used to develop two prototype search interfaces for a mobile device. User feedback was then obtained using the standard SUMI instrument, and both interfaces were confirmed as usable. In contrast to what was expected, users preferred the text interface in general, and according to most SUMI subscales. This could be because of greater familiarity with text search interfaces or because of the relative scarcity of voice interfaces in African (Bantu) languages. Where users are not literate, the voice interface may be the only option, so the fact that it was deemed usable is an important independent finding.

Search in African language collections is still a largely unexplored field, and more work needs to be done on the interfaces as the algorithms and collections are developed in parallel.

Acknowledgment

This research was partially funded by the National Research Foundation of South Africa (Grant numbers: 85470, 88209 and 105862) and University of Cape Town. The authors acknowledge that opinions, findings and conclusions or recommendations expressed in this publication are that of the authors, and that the NRF accepts no liability whatsoever in this regard.

[3]http://sumi.ucc.ie/sumipapp.html#sumidev

References

[1] ITU, "The World in 2014 - ICT Facts and Figures," *International Telecommunication Union*, no. in 2008, 2014. [Online]. Available: http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2014-e.pdf

[2] M. Satyanarayanan, "Mobile information access: A study of Emerging Search Behavior on the Mobile Internet," *IEEE Personal Communications*, vol. 1, no. 1, pp. 1–38, 2007.

[3] M. D. Dunlop and S. a. Brewster, "The Challenge Of Mobile Devices For Human Computer Interaction," *Mobile HCI 01*, vol. 6, no. 4, pp. 235–236, 2002.

[4] M. Chau, X. Fang, and C. C. Yang, "Web searching in Chinese: A study of a search engine in Hong Kong," *Journal of the American Society for Information Science and Technology*, vol. 58, no. 7, pp. 1044–1054, may 2007. [Online]. Available: http://doi.wiley.com/10.1002/asi.20592

[5] I. M. Thies, "User Interface Design for Low-literate and Novice Users : Past , Present and Future," 2015.

[6] M. Hattab, B. Haddad, M. Yaseen, A. Duraidi, and A. Shmais, "Addaall Arabic Search Engine: Improving Search based on Combination of Morphological Analysis and Generation Considering Semantic Patterns," *International Journal*, pp. 159–162, 2007. [Online]. Available: http://www.pu.edu.jo/download/cv/Addlaall-Search-Engine–Hattab-Haddad-Yaseen-UOP.pdf

[7] B.-y. Ricardo, G. Dupret, and J. Velasco, "A study of mobile search queries in Japan," *WWW'07 Workshop*, 2007. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.68.6279

[8] M. Schuster and K. Nakajima, "Japanese and korean voice search," *International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 5149–5152, 2012.

[9] J. Shan, G. Wu, Z. Hu, and X. Tang, "Search by Voice in Mandarin Chinese," *Interspeech 2010*, no. 1, pp. 354–357, 2010. [Online]. Available: http://www.isca-speech.org/archive/interspeech_2010/i10_0354.html

[10] Y. Sung, M. Jansche, and P. Moreno, "Deploying Google Search by Voice in Cantonese," *InterSpeech2011*, no. August, pp. 2865–2868, 2011. [Online]. Available: http://research.google.com/pubs/archive/37116.pdf

[11] E. Barnard, J. Schalkwyk, C. van Heerden, and P. J. Moreno, "Voice search for development," *Interspeech 2010, 11th Annual Conference of the International Speach Communication Association*, no. September, pp. 282–285, 2010. [Online]. Available: http://www.isca-speech.org/archive/interspeech_2010/i10_0282.html

[12] J. Nielsen, "Iterative User-Interface Design," *IEEE Computer Society Press*, vol. 26, no. 11, pp. 32–41, 1993. [Online]. Available: https://dl.acm.org/citation.cfm?id=619982

[13] B. Suhm, "Towards Best Practices for Speech User Interface Design," *Conference on Speech Communication and Technology (Eurospeech)*, vol. IV, 2003.

[14] J. Nielsen, "10 Heuristics for User Interface Design," *Jakob Nielsen's Alertbox*, 1995. [Online]. Available: http://www.nngroup.com/articles/ten-usability-heuristics/

[15] N. Yankelovich and J. Lai, "Designing speech user interfaces," *CHI 98 conference summary on Human factors in computing systems - CHI '98*, no. April, pp. 131–132, 1998.

[16] M. A. Hearst, *Search User Interfaces*, 1st ed. Cambridge University Press, 2009. [Online]. Available: http://searchuserinterfaces.com/

[17] N. Yankelovich, G.-A. L. Levow, and M. Marx, "Designing SpeechActs: Issues in speech user interfaces," *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 369–376, 1995.

[18] J. Kirakowski and M. Corbett, *SUMI: The software usability measurement inventory*. John Wiley & Sons, Inc., 1993, vol. 24. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8535.1993.tb00076.x/abstract

[19] P. W. Jordan, B. Thomas, I. L. McClelland, and B. Weerdmeester, *Usability Evaluation In Industry*. London: Taylor & Francis, 1996.

[20] H. Hsu, P. A. Lachenbruch, H. Hsu, and P. A. Lachenbruch, "Paired t-Test," in *Wiley Encyclopedia of Clinical Trials*. Hoboken, NJ, USA: John Wiley & Sons, Inc., sep 2008. [Online]. Available: http://doi.wiley.com/10.1002/9780471462422.eoct969