THE UTILITY OF SPOKEN DIALOG SYSTEMS

Etienne Barnard, Madelaine Plauché, Marelie Davel

Human Language Technologies Research Group, CSIR Meraka Institute, South Africa
ebarnard@csir.co.za, mad@brainhotel.org, mdavel@csir.co.za

ABSTRACT

The commercial successes of spoken dialog systems in the developed world provide encouragement for their use in the developing world, where speech could play a role in the dissemination of relevant information in local languages. We investigate the evolution of spoken dialog system research in the developed world, and show that the utility of speech is based on user factors and application factors (amongst others). After adjusting the factors for the developing world context and plotting their interactions, we offer several predictions for the field. In particular, we show that the field of spoken dialog systems for the developing world is in a nascent stage and will likely take another decade to have an impact similar to that in the developed world.

Index Terms— *Speech recognition, developing nations, user interface human factors, user interfaces, spoken dialogue systems.*

1. INTRODUCTION

There is a widespread belief that spoken dialog systems (SDSs) will have a significant impact in the developing world [1]. Firstly, speech-based access to information may enable illiterate or semi-literate people, 98% of whom live in the developing world [2], to participate in the information age. Also, the availability of traditional computer infrastructure is low in the developing world, but telephone networks (especially cellular networks) are spreading rapidly. A strong oral culture exists in many traditional societies, which is likely to render such systems more acceptable than text-based or graphical information sources. Finally, the availability of alternative information sources is often low in the developing world.

Based on this perceived value of SDSs in the developing world, a number of exploratory studies have been performed in recent years. Barnard *et al*. [3] report on preliminary experiments performed to assess the usability of a telephone-based information service for access to government information in South Africa. A kiosk-based SDS for agricultural information was developed by Plauché *et al*. [4], and evaluated by semi-literate users in rural Tamil Nadu, India. Nasfors [5] also developed an agricultural information service which is aimed at mobile telephone users and deployed in Kenya. The two most sophisticated SDSs in this category were the telephone-based information service for community health workers piloted in Pakistan [6] and the similar system for caregivers of children with HIV piloted in Botswana [7].

Each of these pioneering studies was primarily aimed at assessing the feasibility of using speech technology in various settings in the developing world. However, in the process of determining feasibility, a number of practical lessons were also learned. For example, it was found that user acceptance of such systems is proportional to the difficulty that users would have to access the same information through other mechanisms [4] (thereby confirming the concept of the "motivated user"), and two studies [3,6] found that it was preferable to use more verbose, less efficient user interfaces to guide inexperienced users for whom time pressure is not a primary concern.

However, these findings are still very preliminary; to gain more insight on the utility of speech in the developing world, we investigate the much more plentiful research on this topic that has been performed in the developed world. Of course, scientific studies of SDS in the developed world assume a context that is different from that occurring in the developing world. Much of the previous design recommendations and usability methods are therefore not directly applicable, and some care must be taken to perform appropriate translations between the two environments.

In this paper, we show how speech R&D in the developed world progressed from descriptive studies to rigorous scientific research (Section 2). We then extract from this trend the relevant user and application factors that determine the utility of speech (Section 3). Finally, we situate the current status of developing world research in this continuum to show that another decade of exploratory research lies ahead before we can expect a rigorous field to emerge. We also present several hypotheses regarding the factors that influence the utility of speech in the developing world, focusing on the interaction between user ability and application complexity (Section 4). We conclude with suggestions for future research in this area to develop SDSs that will most impact the developing world (Section 5).

2. EVOLUTION IN THE DEVELOPED WORLD

In the developed world, spoken dialog systems have come a long way from the digit-recognition systems trialed at Bell Laboratories in 1952 to the current voice portals that are used by millions of callers per day. This evolution rarely followed the classical progression from scientific breakthrough to application in engineering technology. It was much more common for numerous applications to be attempted in somewhat haphazard fashion, with little reliance on rigorous scientific foundations. Some of the application areas where SDSs had been expected to play a major role never materialized, whereas unexpected successes for SDSs arose elsewhere. We now illustrate this process by reviewing the history of call-center automation in the developed world.

2.1. Case Study: Call-Centers

Some of the earliest commercially successful applications of spoken dialog systems were in answering telephone calls that would otherwise be handled by human operators, typically in call-center environments. Such "call-center" applications were easily seen to be economically important: as speech technology was maturing in the last decade of the twentieth century, the call-center industry had annual revenues of several billion dollars; the industry was expanding rapidly and suffered from significant personnel turnover. Spoken dialog systems were seen to offer a number of potential advantages in this environment [8, p10], including the following:

Cost savings. SDSs are significantly less expensive to maintain than human operators, but able to achieve better automation rates than competing technologies such as those based on telephone key presses ("DTMF systems").

Branding. *S*ince the spoken modality is deeply ingrained in the human mind, spoken communication creates associations that can be employed to extend and enhance the brand of a company. In this respect, the repeatability and control offered by an automated system are particularly attractive in an industry where burnout and turnover of personnel are significant factors.

Efficiency. Well-designed SDSs make it possible to exchange information in an efficient manner, thereby cutting down on call durations, reducing toll costs and (hopefully) increasing customer satisfaction.

The benefits of an SDS relative to DTMF systems are most significant when the information to be exchanged is not easily represented on a numerical key pad. Therefore, the early SDS applications for the call-center focused on tasks such as directory assistance [9], stock quotes [8, p3] and travel information systems [10,11], which require selections from long list of names of people, companies or locations. Since little research was available on the best way to design interfaces, much of the early development was based on a combination of intuitive insights and informal focus groups. (Interestingly, this "pre-scientific" investigation of the space of possible applications is quite similar to the current state of affairs for developing-world applications.) In this way, a wide range of applications were shown to be commercially viable, and today SDSs are used in thousands of call-center applications [8, p9] as well as derived services such as voice portals and automated personal assistants.

During this process of informal development, several false starts and surprising failures were encountered. For example, there was much early enthusiasm about self-service applications that allow customers in large industries to order and configure their paid services. In the telecommunications industry, a leading prospective adopter of such services, users would be allowed to configure services such as call waiting, call forwarding and answering services (see, for example [12]). Several trial systems were developed, and performed well in pilot tests. However, (in contrast to, for example, call routing services) such self-service applications were never a major commercial success, and it was generally found that companies preferred to continue offering such services through human operators. The self-service applications were shown to function well from an end user's perspective; the main reason for their limited application was apparently economic - human operators are able to create up-selling opportunities in a way that automated systems cannot do.

Similarly, the branding advantages of services enhanced with SDSs were found to be much smaller than had been expected [13]. Whereas technologists tend to be biased by the technical sophistication of spoken interaction, consumers are more concerned with the limitations that cause fragility and unexpected behavior. As a consequence, a survey reported in [13] found that 45% of consumers prefer to have the automated spoken component in these applications limited to a minimum, compared with 9% of vendors who responded to the survey.

After more than 15 years of intensive development in this application domain, the technical as well as economic factors that determine the viability of SDSs in call-center applications are now fairly clear. Major progress in the understanding of the technical factors that determine user acceptance followed from the development of reliable instruments for assessing user experiences [14] along with the systematic development and use of associated protocols for administering and assessing such instruments [15]. Target users were included in the design process and usability methods employed both during design and analysis [8]. Thus, SDSs are now widely used in routine information services such as stock quotes, package tracking and confirmation of flight arrival and departure times, whereas the associated transactional services (reserving the flights, purchasing or selling the shares) are more commonly handled by human operators or other modalities such as Web interfaces. Much has been learned about the user populations that are most receptive for these applications (they tend to be youthful, technologically sophisticated and first-language speakers of the language in which the SDS operates [16]). A diverse industry of hardware vendors, software vendors and various service specialists has grown to support such services (see, for example, the wide range of companies listed at http://www.speechtechmag.com/ VerticalMarkets/). These companies are able to scope, design and deliver viable commercially-oriented solutions in the developed world with a high degree of predictability.

3. FACTORS IN UTILITY OF SPEECH

The deemed utility of speech as a mode of interaction with technology depends on factors including but not limited to cost, availability of training for the user, payoff, complexity of the application, user needs, and as previously discussed, quality of the design and development process. In this section, we will focus only on the factors that relate to the application (Section 3.1.) and those that relate to the users (Section 3.2.). We will draw primarily from developed world research, such as the case study previously mentioned, but we will adjust the factors for developing world environments.

3.1. Application Factors

SDS applications in the developed world have revealed several factors that influence the utility of speech as an input modality. Application factors include the following:

- **Complexity**. This includes the restrictiveness of the task domain, the linearity of the interaction required and the range of choices available.
- **Capability.** This includes the speed, accuracy and robustness of the speech recognition that powers an application, specifically with regard to noise and non-standard speech.

- **Design.** This includes hierarchical versus flexible navigation, choice of input style (such as command-based, grammar-based or natural language), application of speech user interface design best practice (such as summarized in [17]) and general interaction style.
- **Environment.** This includes the possible need for hands-busy or eyes-busy system interaction and the need for privacy, with the former a strong indicator for SDS and the latter a strong counter-indicator.
- **Alternative solutions.** Whether alternative means of system or information access exist can also contribute to user motivation.

Research studies that focus specifically on the interplay between user and application (mainly related to SDS preference over DTMF systems) have found that the type of task has a strong influence on a user's ability to conclude a task effectively and efficiently. (These influences are expected to be somewhat different in the developing world, as we discuss below.) Speech input fares better for complex (non-linear, unrestricted, broad domain) tasks and DTMF is more effective when tasks are simpler (linear, limited options) [17]. Various studies have found a discrepancy between user performance and user preference when evaluating both in laboratory studies [17], even though this tendency is not expected to hold true during continuous, long-term usage of systems.

3.2. User Factors

User factors that have been considered in the literature of the developed world as relevant to the utility of speech include age, gender, cultural background, income levels, spatial and/or verbal ability, disability, education and experience in using similar systems, and whether the effort involved in learning to use a system has adequate payoff. User factors may also influence the concept of the "motivated user", where intrinsic user characteristics (such as disability) prevent a user from interacting with systems except through an SDS.

Several "user ability" factors are particularly salient for SDS deployment in the developing world, including:

- **Literacy.** Target populations of many of the case studies include illiterate or semi-literate users. This factor is closely linked to education.
- **Experience.** Target populations often contain users who have never been exposed to similar technologies.
- **Access to Training.** Training for new technologies is often limited in developing regions.
- **Access to Devices.** A more limited selection of devices applies to target users (e.g. environments where only basic telephony handsets or kiosks are available, rather than a choice of devices).
- **Income.** Services that require even the cost of a local phone call can be prohibitively expensive for many individuals in developing regions.
- **Language.** More heterogeneity of languages and dialects is present.

Intrinsic user factors are usually not sufficient to predict a user's preference for a particular modality; rather, it is the interplay between user and application that influences a user's preference. The utility of speech, in turn, has some correlation with user preferences [17] as well as other factors.

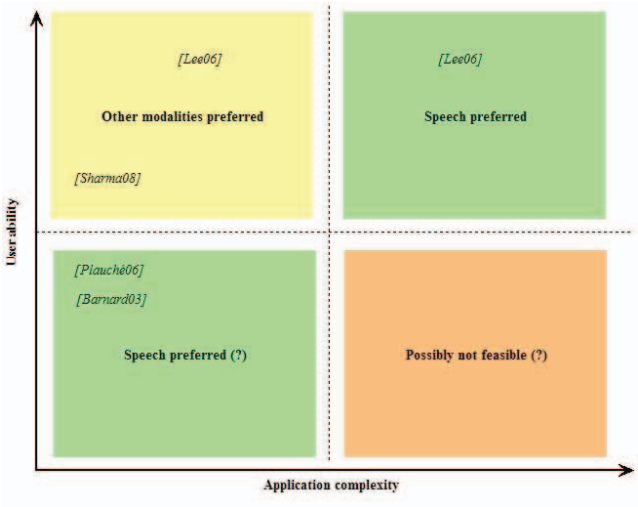

Figure 1: The preference for spoken interaction is expected to depend on an interaction between user sophistication and application complexity

4. PREDICTIONS FOR THE DEVELOPING WORLD

Research related to successful and sustainable deployment of SDS in the developing world is unfortunately fairly rare, with many initial studies not having progressed beyond a pilot phase. In addition, the case studies of Section 2 show that extensive real-world deployment may be necessary before we can arrive at robust results on the acceptance and impact of such systems. However, the relevant factors for the utility of speech determined from research in the developed world, when adjusted to developing world conditions, may provide a framework from which we can bootstrap predictions and evaluations of the impact of speech systems in this new domain. Here, we focus on factors related to user ability and application complexity since these are likely to be major determinants in the developing world.

We group the following user factors under the umbrella term, *User Ability*: Literacy, Experience, Access to Training. We also consider the application factor *Application Complexity,* which refers to factors such as the restrictiveness of the task domain, the linearity of the interaction required and the range of choices available. We can examine the interactions between User Ability and Application Complexity and can summarize the differences between the two environments compactly. While these are by far not the only factors of importance, these two sets of factors can be used to sketch a framework of the space in which research is currently required (Figure 1).

Given the current state of the art in speech recognition and SDS design, spoken interfaces have significant limitations compared to human abilities. For users scoring high on User Ability, developed-world studies indicate that Application Complexity strongly influences the preferred modality. Such users find that simple tasks are more easily performed using non-speech modalities (such as DTMF for IVR systems, or touch screens at kiosks) and more complex tasks ("How Can I Help You?") better executed using the more flexible medium of speech. However, when users score low

with regard to User Ability, the expectation is that speech becomes the preferred medium, even for very simple tasks. Whether complex tasks can be executed in this domain (high application complexity, low user ability) has not yet been established. Note that User Ability includes factors such as exposure to similar systems and training received (or experience gained) with the SDS of interest. As the ability of a user (from either the developed or developing world environment) increases, it is expected that modality preferences will change.

Our analysis of the interaction between the selected user and application factors raises three hypotheses, as suggested by the illustration in Figure 1:

Hypothesis 1: Naïve users in the developing world will prefer the speech modality for simple applications which are generally best implemented with other modalities in the developed world.

Hypothesis 2: Through focused user training (or extensive experience with an available SDS), developing-world users will take on preferences more similar to those observed in the developed world.

Hypothesis 3: In the absence of training or exposure, there is a large class of applications that will not be feasible in the developing world; those applications will – somewhat surprisingly – share many of the features (non-linearity, diversity of user choices) that render speech the preferred modality in the developed world.

Further predictions based on lessons learned from the developing world context include the need to employ usability methods or similar evaluation procedures to assess user experience. Even despite quality design and evaluation procedures, some tasks will continue to require human rather than machine interaction and cost will always be a driving factor in speech-based systems.

5. CONCLUSION

As the use of spoken dialog systems becomes increasingly widespread in the developed world, it is easy to forget the trial-and-error development that led to the rigorous design and evaluation methods used today. If a similar trajectory is required for a comparable impact of SDSs in the developing world, we should expect a decade or more of pilot deployments of such systems as speech technologists grapple with issues of feasibility, sustainability and user acceptance.

It is nevertheless possible to identify similarities and differences between systems and users in both the developed and developing world contexts. It is also becoming clear what questions must be studied more explicitly. For example, there is still very little evidence that complex applications will be suitable for less experienced, less literate users in the developing world.

There is no doubt that a scarcity of relevant, up-to-date information sources is one of the gravest deficiencies of the developing world, and we remain convinced that spoken language technologies can play a significant role in addressing this issue. Hopefully, a careful analysis of the lessons learned in the developed world, along with an understanding of the salient differences in the developing world, will guide the research community towards the performance of the appropriate trials and experiments that will accelerate speech to reach its potential as a universal and accessible means for information dissemination in the developing world.

6. REFERENCES

[1] R. Tucker and K. Shalonova. "The Local Language Speech Technology Initiative", SCALLA Conference, Nepal, 2004.

[2] Summer Institute of Linguistics, “Facts about Illiteracy,” Online: http://www.sil.org/literacy/litfacts.htm, accessed June 2008.

[3] E. Barnard, L. Cloete, and H. Patel. “Language and Technology Literacy Barriers to Accessing Government Services,” Lecture Notes in Computer Science, vol. 2739, pp. 37-42, 2003.

[4] M. Plauché, U. Nallasamy, J. Pal, C. Wooters, and D. Ramachandran. “Speech Recognition for Illiterate Access to Information and Technology,” In Proc. IEEE Int. Conf. on ICTD, pp. 83-92, May 2006.

[5] P. Nasfors. "Efficient Voice Information Services for Developing Countries", Master Thesis, Department of Information Technology, Uppsala University, Sweden, 2007.

[6] J. Sherwani, N Ali, S. Mirza, A. Fatma, Y. Memon, M. Karim, R. Tongia and R. Rosenfeld, “Healthline: Speech-based Access to Health Information by low-literate users”, In Proc. IEEE Int. Conf. on ICTD, Bangalore, India, Dec. 2007.

[7] A. Sharma, M. Plauché, C. Kuun, E. Barnard, “HIV Health Information Access using Spoken Dialogue Systems: Touchtone vs. Speech,” submitted to IEEE Int. Conf. on ICTD.

[8] M.H. Cohen, J.P. Giangola, and J. Balogh, “Voice User Interface Design”, Addison-Wesley, Boston 2004.

[9] M. Lennig, G. Bielby and J. Massicotte, “Directory assistance automation in Bell Canada: Trial results”, Speech Communication, vol. 17, no 3-4, pp 227-234, Nov 1995.

[10] E. Barnard, A. Halberstadt, C. Kotelly and M. Phillips, “A consistent approach to designing spoken-dialog systems”, ASRU Workshop, Keystone, Colorado, 1999.

[11] H. Strik, A. Russel, H. Van den Heuvel and C. Cucchiarini, “A spoken dialog system for the Dutch public transport information service”, International Journal of Speech Technology, vol. 2, no 2, pp. 121-131, 1997.

[12] D. Johnston, “Telephony based speech technology - from laboratory visions to customer applications”, International Journal of Speech Technology, vol. 2, no 1, pp. 89-99, 1997.

[13] C Musico, “Press 1 for caller thoughts”, Speech Technology Magazine, Aug. 2008.

[14] KS Hone and R Graham, “Towards a tool for the Subjective Assessment of Speech System Interfaces (SASSI)”, Natural Language Engineering, vol 6, no 3–4, pp 287–303, 2001.

[15] B Suhm, “IVR usability engineering using guidelines and analyses of end-to-end calls”, in D. Gardner-Bonneau and HE Blanchard, eds. “Human factors and voice interactive systems”, Springer, 2008.

[16] R. Pieraccini and D. Lubensky, “Spoken language communication with machines: the long and winding road from research to business”, Lecture Notes in Computer Science, vol. 3533, pp. 6-15, Springer-Verlag, 2005.

[17] K.M. Lee and J. Lai, “Speech Versus Touch: A Comparative Study of the Use of Speech and DTMF Keypad for Navigation”, International Journal of Human-Computer Interaction, vol. 19, no 3, pp. 343-360, 2006.