The Case for Speech Technology for Developing Regions

J Sherwani and Roni Rosenfeld
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA
{jsherwan,roni}@cs.cmu.edu

ABSTRACT

Many ICT initiatives that attempt to connect end-users to each other rely on Internet-connected PCs using graphical user interfaces (GUIs). While speech-based systems over the telephone have mushroomed in the West in the past decade, their potential remains largely untapped for use in the developing world. We present reasons why speech systems may be preferable to traditional PC/GUI systems in these circumstances, along with the needs such systems can fulfill.

1. INTRODUCTION

The role of ICTs in sustainable development has been clearly documented [1], and this enterprise has recently been coined “ICT4D” (ICTs for Development). The form that many ICT4D initiatives take is that of PCs using GUIs connected to the Internet. Systems using spoken language as a modality, operating over standard telephony devices, however, offer a cheaper and perhaps more usable alternative. Such systems are widely deployed as customer self-service applications in the West – for tasks such as travel information and reservations, movie showtimes and tickets purchasing, and call routing. Such systems are even more compelling for low literate users in developing countries, although their use in such contexts has not been deeply explored.

In this paper, we present the case for SLT4D (Speech and Language Technologies for Development). In the next section we examine the reasons that make speech compelling in the context of under-served populations. In section III we discuss what niches speech technology can serve in this context. Finally, we dispel some common misconceptions about speech technology.

2. WHY IS SPEECH COMPELLING?

Speech-based systems can be compared to traditional ICT initiatives using Internet-connected PCs, which provide a reasonable baseline for comparison. Such systems typically require:

1. Deployment of PCs in underserved communities
2. Internet connectivity (broadband for multimedia)
3. Literate users
4. Users willing to learn how to use computers
5. Languages with written forms
6. System maintenance and upgrades at the user end
7. Protection against accidental misuse of the PC[1]
8. Protection against computer viruses

Telephony-accessible speech-based systems compare quite favorably, and have the following requirements:

1. Deployment of telephony devices in underserved community
2. Landline or cellular connectivity
3. Literacy is not required
4. Interaction is conversational – less user adaptation is needed
5. Any spoken language can be supported, including switching among languages in the same session
6. System maintained and upgraded centrally
7. No protection necessary against accidental misuse
8. No protection necessary against viruses

Finally, rates of telephony growth in developing countries dwarf rates of PC- and Internet-penetration. Models such as GrameenPhone [4] indicate that there are viable, sustainable mechanisms of telephony use in underserved communities, while similar models for sustainable PC & Internet use require much higher cross subsidization [2].

Thus, we believe there is a strong advantage, from both sustainability and human factors perspectives, for telephony-based speech applications to be favored over PC-based ICT initiatives. We now turn to a discussion of what needs speech technologies can fulfill.

Conference’04, Month 1–2, 2004, City, State, Country.

[1] In the e-Choupal initiative [2], one recurring problem was the inadvertent deletion of the desktop shortcuts from kiosks by users.

3. SPEECH APPLICATIONS FOR LOW LITERATE USERS

Interaction with a speech system begins with either the user calling in the system or vice versa. Once the channel has been established, in principle any transaction that could be carried out on a web-page can be handled through telephony, although some types of transactions may be more amenable to speech interfaces than others. Here we present a list of sample applications that is in no way exhaustive:

Information access: weather information, crop market price information, health information, government records, Orbitz-like aggregation for market prices.

Information entry: sellers entering price and quantity of goods, lodging service complaints, disaster information entry by survivors.

Information management: local content where users contribute information and can look it up later – can be collaborative, shared access, or private – hence an audio equivalent of Windows Notepad, blogs, message boards, and wikis, which enables virtual community formation as well as the grass-roots creation of shared cultural repositories.

Transactional needs: reservations, sales & purchases.

Verifiable proof mechanisms for low-literate users: proof that a transaction has been completed, that an agreement has been reached, that a request has been filed, or that payment has been made (conventional paper-based alternatives are not useful for low-literates).

Given the above possibilities, it is surprising that there aren't many more speech-based systems in current ICT initiatives. We feel this may be a result of a few misconceptions about speech systems for ICTs, which we now briefly discuss.

4. SPEECH INTERFACE MYTHS

Myth #1: Until speech recognition is 100% accurate, speech systems aren't useable. Even human-human conversations are error-prone. However, task success can be ensured through the judicial use of confirmations, error-recovery dialogs, and other dialog mechanisms. This is analogous to end-to-end error correction in network communications – while any specific turn in the conversation may be erroneous, the entire communication can be guaranteed to be successful with a high degree of certainty. This is especially true when the users are highly motivated, perhaps because they have no easier alternatives, as is often the case with under-served or low-literate communities.

Myth #2: Building speech technologies for new languages is prohibitively expensive. Building a baseline speech synthesis system is now so well understood that it is routinely used as a student assignment in graduate courses on speech synthesis. Baseline speech recognition systems can also be created with modest time investments [6]. Additionally, there now exist tools and technologies for rapidly building capacities in new languages [3].

Myth #3: Populations in developing countries have many dialects and accents, making it impossible for a system to recognize their speech accurately. There is nothing intrinsically unrecognizable about a specific accent – what matters is how different the speech training data is from the speech of the actual user population. Through bootstrapping with an initial system, more speech data can be gathered which can be used to better train the system to work with the target population's accent.

5. CONCLUSION

Speech interfaces can prove revolutionary in the context of the developing world. Given the number of projects investigating cell phone-based services in this context (such as [5]), it is surprising that there are not many projects investigating spoken language interfaces, since these have lower requirements in literacy as well as in device cost. Initial research on such speech interfaces [6, 7] shows great promise, and we hope that other research initiatives begin to explore this field.

6. ACKNOWLEDGMENTS

We would like to thank Microsoft Research for funding the HealthLine project.

7. REFERENCES

[1] Brewer, E., Demmer, M., Du, B., Fall, K., Ho, M., Kam, M., Nedevschi, S., Pal, J., Patra, R., and Surana, S. "The Case for Technology for Developing Regions". IEEE Computer. Volume 38, Number 6, pp. 25-38, June 2005.

[2] Annamalai, K & Rao, S, "What Works: ITC's eChoupal and Profitable Rural Transformation", Digital Dividend Business Case Study, August 2003, http://www.digitaldividend.org/pdf/echoupal_case.pdf

[3] Black, A. & Schultz, T., "SPICE: Speech Processing – Interactive Creation and Evaluation Toolkit for New Languages", http://www.cmuspice.org

[4] Cohen, N., "What Works: Grameen Telecom's Village Phones", Digital Dividend Business Case Study, June 2001, http://www.digitaldividend.org/pdf/grameen.pdf

[5] Parikh, T. "Using Mobile Phones for Secure, Distributed Document Processing in the Developing World". IEEE Pervasive Computing Magazine, 4(2):74–81, April 2005.

[6] Plauche, M., Nallasamy, U., Pal, J., Wooters, C., and Ramachandran, D. "Speech Recognition for Illiterate Access to Information and Technology". Proc. Information & Communications Technologies and Development, 2006.

[7] Sherwani, J., Ali, N., Mirza, S., Fatma, A., Memon, Y., Karim, M., Tongia, R., Rosenfeld, R. "HealthLine: Speech-based Access to Health Information by Low-literate Users". In Proc. Information & Communication Technologies for Development, Bangalore, India, December 2007.