The Speech-Enhanced World

Botond Pakucs

CTT, Centre for Speech Technology, Dept. of Speech, Music and Hearing
KTH, Royal Institute of Technology,
Drottning Kristinas väg 31, 100 44 Stockholm, Sweden
`botte@speech.kth.se`

1. Introduction

Speech interfaces are currently about to be integrated in consumer appliances and embedded systems and are expected to be used in mobile and ubiquitous computing environments. However, the new environment and the new interaction modes with appliances and services may increase the traditional requirements on speech based interfaces and may even introduce new human computer interaction and usability related problems.

In spite of the rapid development of the speech technology during the last decades, speech operated services are still seldom used. This is because speech technology is quite expensive and because it is considered rather unreliable and not robust enough for many applications.

So how can we provide high performance solutions that facilitate immediate user success and are suitable for the targeted market's tight cost constraints?

2. The proposal

2.1. Personalized and generic speech interfaces

We suggest a user-centered, application independent model for speech interfaces for mobile and ubiquitous computing environments, where every user is expected to use a *SINGLE*, highly personalized speech interface to access a multitude of services and appliances. We believe that a user-centered *single user and multiple application* speech interface is an appropriate way to provide widely available user-friendly speech services.

This solution relies on and also supports personalization, adaptivity, context awareness and user modeling. The user-centered speech interface could be integrated into some personal, wearable appliance such as a mobile phone or a PDA. In such a case, the speech interface would always be accessible and with all user-dependent data constantly activated and ready to use.

2.2. The Speech-Enhanced World

Service descriptions, namely domain dependent data, such as knowledge-models, dialogue descriptions could be developed similarly to how VoiceXML based telephony services are developed today. These service descriptions should be stored locally on the service provider side, i.e. in the appliances, and loaded dynamically into the personalized speech interface when they are needed in the appropriate context and environment.

In addition to the service descriptions some storing capacity is all that is needed together with a wireless communication capability on the service provider side. Thus, it would be fairly easy to develop and provide speech interfaces for a wide range of appliances and services, from door-knobs to even spaceships.

2.3. Some major advantages

The user-centered model offers a more cost effective, computationally less demanding, simple to develop and maintain solution for the *service providers and the industry.*

This model makes it unnecessary for *users* to adapt to, and to train several different interfaces. Due to the user's familiarity with the speech interface, user satisfaction will be quite reasonable.

The user-centered model would also make it easier for the *speech technology community* to develop robust and high performance speech interfaces. This solution makes it possible to use speaker dependent speech recognition. In this way the speech recognition error-rate could be decreased substantially. All services and appliances can be made available to all users, in spite of dialects, non-native accents and even speaking disorders, through adapting the speech recognition to individual users. This solution is also ideal to gather data on the user's behavior and speech patterns. This data could be used to build user models, which could be useful in dialogue management to predict user intent given the current dialogue context.

2.4. Feasibility

An user-centric speech interface solution requires an advanced infrastructure. However wireless communication, ad-hoc networking, ubiquitous computing etc. are very active research areas and thus, not a major obstacle.

The major challenge is to build a generic speech interface with support for adaptivity, personalization, context aware computing and with seamless access to diverse applications. The first step towards such an interface is to build a generic dialogue-manager for mobile and ubiquitous computing environments. We believe that this is feasible. This is what my research is about [1].

3. Discussion

Making a multitude of services and appliances available to everyone by means of the voice is a wonderful vision. We believe that the user-centric approach has some major advantages and is promising enough to be worthy of further research and evaluation.

4. References

[1] Pakucs, B. "User-Centered Dialogue Management for Mobile and Ubiquitous Computing Environments". To be published, 2001.