Developing Speech Input for Virtual Applications: A Human Factors Perspective

Alex Stedmon, University of Nottingham, UK
David Howard, University of York, UK
Christin Kirchhübel, University of York, UK

ABSTRACT

This paper contextualises the position of speech input from a user-centred human factors perspective. It is presented as a position paper so that researchers and designers can consider the underlying and future factors of a people-orientated approach to speech input for virtual applications. A number of key areas are explored including: human factors for speech input; speech input for virtual applications; speech as a spare mode of interaction; user acceptance and uptake; incorporating speech in the development of virtual applications; and speech input as an interaction tool. Given the user-centred perspective of this paper; this paper does not set out to address issues associated with spoken dialogue technologies, dialogue, and dialogue management; recent work on conversational agents in virtual environments; or multimodal interaction. This paper places the focus more fundamentally within human factors by looking at the user first as a basis for developing usable virtual applications incorporating speech input rather than to review the current state of the art in interaction design. A particular point this paper makes, however, is that speech input should be designed and used as another interaction tool that users need to learn to use, rather than assuming it will offer a natural or intuitive interface.

Keywords: Interaction Devices, Speech Input, Speech Recognition, User-Centred Design, Visual Applications

1. INTRODUCTION

In 1952 Bell Labs introduced one of the first speech recognisers which achieved 97% accuracy (Davis, Biddulph, & Balashek, 1952). Original systems were limited by their capacity to only recognise isolated words and small vocabulary sets; however, this input modality appeared to offer a ‘natural’ mode for human-machine communication that if attainable in a cost-effective way, would be unsurpassed in making computers cooperative systems, rather than increasing the demands on the user to adapt to the machine (Lea, 1980). Since then, speech recognition has been perceived as a natural interface, but it is unclear whether the development of speech recognition technology is aimed “solely at the technology or at the user’s interaction with the system” (Baber & Noyes, 1996, p. 149). If speech recognition evolves along a purely technological route then...
there is a danger of designing systems that do not support user needs or expectations, and if systems evolve along purely interaction based principles then there is a danger of not embracing or exploiting new technological capabilities to their full.

To date speech recognition has yet to be implemented in an original application (that has not used another form of input device before) and, as a consequence, it always faces the immediate challenge from more conventional input devices (e.g., keyboard and mouse) that users are familiar with using. Furthermore, if user expectations of speech input are too high, there is a danger that its potential will not be realised due to user frustration leading to poor uptake. The increased availability of speech input “brings with it the need for a full understanding of the ergonomics aspects of these systems with the aim of developing general guidelines to ensure their application should be as effective as possible” (Hapeshi & Jones, 1988, p. 252).

Whilst the potential to interact with machines using speech input has been possible for many decades (Ullman, 1987) and technical advances, especially with the development of computing and distributed interaction technologies, allow for new approaches to people orientated interaction, speech input still remains an elusive concept without widespread use or end user acceptance (Stedmon et al., 2011). Alongside the point that user expectations might be higher than the technology can deliver, system interaction is often not as intuitive as it could be. With a lack of consideration for user needs, user satisfaction and the adoption of new technology might be hampered by fundamentally unusable technologies. It is only when system developers and users are familiar with the underlying issues of development and use, that such technologies can be designed that will embrace user needs and technological ability in a balanced manner.

With a focus on design needs for user-centred and usable speech interfaces, along with a framework for virtual environment (VE) development, this paper argues that there has been little recent development in understanding generic human factors issues of speech input and even less in the specific area of human factors for speech input in virtual applications. Given the user-centred perspective of this paper and the aim of presenting arguments that transcend specific technologies or trends in solutions, this paper does not set out to address issues associated with natural and spoken dialogue technologies (Cohen, Giangola, & Bagola, 2004; Leuski & Traum, 2010), dialogue and dialogue management (Lemon, 2011) or recent work on embodied conversational agents in virtual applications (Traum, 2008). Furthermore, the technologies and methods underpinning multimodal interaction (Lee & Billinghurst, 2008), tangible or mobile interfaces (Billinghurst, Kato, & Myojin, 2009) is not a primary focus of this paper either. And whilst some researchers may consider it difficult to address the nature of speech-based interaction without addressing principles of dialogue and dialogue management, this paper seeks to place the focus more fundamentally within human factors by looking at the user and then seeking to develop usable virtual applications incorporating speech to support user interaction rather than to review the current state of the art in interaction design.

This paper discusses the importance of speech input across a number of key areas:

1. Human factors issues in speech input,
2. The interaction potential of speech input,
3. Acceptance issues of speech input in virtual applications,
4. Designing virtual applications with speech input in mind,
5. Developing speech input for virtual applications,
6. Speech input as an interaction tool.

2. HUMAN FACTORS ISSUES IN SPEECH INPUT

A number of reviews have been conducted that provide a valuable resource for researchers and systems designers alike. From these, human factors of speech input can be seen to have