

2.2 Multimodal and Multichannel UI

One of the breakthroughs users can benefit of is that automatic speech recognition (ASR) improved highly significant over the last years. ASR now works well for dictation tasks. However, dictation is a highly specific use case which does not require the extraction of semantics from the utterances. Some applications use speech input for form filling. However, filling each single slot by speech is often not more efficient than typing.

The question arises: What are important challenges in using speech as a <mainstream= modality? While ASR made significant efforts within the last years, e.g. partly driven by the successful application of deep neural networks, the identification of the intended semantic for a further processing by the dialog manager is still a rather difficult process. ASR capabilities are easy to integrate into new user interfaces by making use of available programming APIs.

On the technical side one of the next challenges is therefore to realize conversational speech interaction in many applications. This requires to simplify the usage of NLP methods for information extraction, dialog processing and presentation, so that developers can easily deploy speech interfaces.

Since the Internet is mobile nowadays and conversational speech is the most convenient interaction mode of complex applications that require more than simple gestures, this will enable even more services at the hand of the users. In that matter it is important to better understand the specific benefits that emerge for individual users. Information about these benefits can be revealed by observing the user's modality choice behavior.

Understanding the factors influencing user's modality choice will enable interface designers to adapt applications to the advantage of the user, and to inform the user about extra possibilities of interaction.