

Spoken Natural Language Technology, Applications and Challenges

Richard V. Cox
AT&T Labs – Research

It is 2002 and no one yet speaks to a HAL 9000-like computer in the same ways that the astronauts did in 2001 – a Space Odyssey. Significant progress has been made in many areas. We are beginning to see the first spoken natural language interfaces for commercial products and services. These interfaces can be purely spoken, e.g. for telephony, or can be multimodal, e.g. for computers or handheld devices. The goals of this tutorial are:

- (1) Provide researchers and developers a perspective of the speech processing technologies that compose voice-enabled and multimodal applications. These technologies include automatic speech recognition, text-to-speech synthesis, natural language understanding, dialogue management, and natural language generation for purely voice-enabled services. For multimodal services, additional technologies include simultaneous spoken and pen-based input (gestures and handwriting recognition) and visual text-to-speech synthesis for output.
- (2) Review the applications and services that already exist today, either in commercial deployment or in laboratory prototypes. Primarily, these examples depend heavily on the context of the application or service, thus lessening the complexity of the task by restricting the domain.
- (3) Address the technical challenges that must next be overcome. These include improvements in the component technologies such as greater robustness in speech recognition and greater naturalness in text-to-speech synthesis, and also include challenges in how to scale the expertise needed to build hundreds and thousands of spoken natural language interfaces for a multitude of businesses.
- (4) Provide a vision of what will be possible in the next five years

The tutorial will include audio and video examples and demonstrations to illustrate the current state of the art.