

Toward Spontaneous Speech Recognition and Understanding

Sadaoki Furui
Tokyo Institute of Technology

Contents

1. Introduction
2. Difficulties of spontaneous speech recognition
3. Acoustic and linguistic models for spontaneous speech recognition
4. Paradigm shift from speech recognition to understanding
5. Message-driven speech recognition and understanding
6. Speech summarization as speech understanding
7. Research projects on spontaneous speech corpus and processing technology
8. Research directions

Abstract

Although high-recognition accuracy can be obtained for speech in the form of reading a written text or similar by using state-of-the-art speech recognition technology, the accuracy is quite poor for freely spoken spontaneous speech. The principal reason for this is that acoustic and linguistic models used in speech recognition have been built using written language or speech reading text, while spontaneous speech and written language considerably differ in both acoustically and linguistically. It is indispensable to build a large corpus of spontaneous speech, since our knowledge of the structure of spontaneous speech is very limited. A paradigm shift from speech recognition to understanding is essential, where underlying messages of the speaker are extracted, instead of transcribing all the spoken words. From these perspectives, several research projects for raising the technological level of spontaneous speech recognition and understanding have recently been conducted in several countries. This tutorial finally predicts future directions in spontaneous speech understanding technologies. It presents the most important research problems, and tries to forecast where progress will be made in the near future.