Abstract

Speech recognition technology offers an attractive interface option: speak to a computer, and it will understand you. One of the most promising applications of speech recognition technology is the spoken dialog system, which offers the promise of simple, direct, hands-free access to information. However, many factors conspire to make user communication with such a system less than optimally efficient. One problem is that users often speak beyond the bounds of what the computer is programmed to understand. This can lead to misunderstandings from the perspectives of both the user and the system, and recovering from such situations can add extra turns and time to the overall interaction. In this thesis, I describe a strategy, termed shaping, for improving user interaction efficiency with spoken dialog systems. This strategy involves the use of a target language designed to foster more efficient communication, and within which users will be encouraged to speak. When users interact with the dialog system and speak outside the target language, the system attempts to understand their input and aims to strike a balance between helping them complete the current task successfully and helping them increase the efficiency of future interactions by learning the target language (which in this case is Speech Graffiti).

The shaping strategies have been investigated through a series of three user studies with telephone-based spoken dialog systems. Results show that shaping can improve efficiency by removing the need for a pre-use tutorial and by reducing word-error rates. Users in the studies exhibited significant intrasession, intersession, and cross-domain increases in Speech Graffiti grammaticality. The studies in this thesis have demonstrated a fully-functional, non-directed-dialog system, accessing real-world data, that takes advantage of users' propensity for convergence.