Chapter 7

Issues in Spoken Dialogue Systems for Human–Computer Interaction

Tanveer J. Siddiqui
University of Allahabad, India

Uma Shanker Tiwary
Indian Institute of Information Technology Allahabad, India

ABSTRACT

Spoken dialogue systems are a step forward towards the realization of human-like interaction with computer-based systems. This chapter focuses on issues related to spoken dialog systems. It presents a general architecture for spoken dialogue systems for human-computer interaction, describes its components, and highlights key research challenges in them. One important variation in the architecture is modeling knowledge as a separate component. This is unlike existing dialogue systems in which knowledge is usually embedded within other components. This separation makes the architecture more general. The chapter also discusses some of the existing evaluation methods for spoken dialogue systems.

1. INTRODUCTION

Ideal communication model for Human-Computer Interaction (HCI) can be derived from human–human interaction which includes both verbal and non-verbal components. Non-verbal communication includes sign languages, facial expressions, gestures, emotions, lip movement, etc., while the main component of verbal communication is a natural language utterances. Understanding human-human interaction process and deriving a computational model of it is an enigmatic problem and achieving an interaction with computers that can be called close to human–like, if not similar, is still too far to achieve. However, achieving human-like interaction has been a desired goal of human computer interaction and the fascinating idea of using natural languages for interacting with computers has long been a research topic of Artificial Intelligence (AI) and HCI researchers.

DOI: 10.4018/978-1-4666-0954-9.ch007
Systems, like ELIZA, which uses natural language interface, have been developed as early as in 60s. Science fictions are full of wonderful pieces of fantasies about spontaneous human-like conversations with computers. The idea of developing “talking computers” (Samuel et al., 2010) has been haunting researchers in AI and speech technology for past few decades. However, spoken language as a means of communication with computers has become a reality only in recent past owing to the rapid advancements in computing, speech and language technology. Several research prototypes as well as commercial applications are now available that use spoken language communication. Interfaces that use voice as input for controlling appliances, creating documents, searching in an existing database, are already in place. However, systems equipped with such voice-based interfaces have limited capability in that they do not engage the user in a natural conversation. They accept speech inputs, process them and perform some actions or report an error. Research efforts are continuing to develop spoken dialogue systems which “provide an interface between a user and computer-based application in many cycles in a relatively natural manner” (McTear, 2002). Figure 1 shows a wide range of tasks of varying complexity suitable for dialogue-based interfaces. However, supporting such interactions brings additional complexities, such as interpreting the user input to get involved in a conversation, handling uncertainty, knowing its state, confirming, asking to repeat, etc.

In order to have more effective interaction dialogue systems must integrate speech with other natural modalities such as facial expressions, gestures, eye movements, etc. This has lead to the idea of Embodied Conversational Agents (ECA) (Traum & Rickel, 2002) either in the form of animated humanoid avatars or, as talking heads. These agents communicate with the user to plan a task that an agent is able to carry out independently, e.g., book travel tickets, prepare diet chart to improve poor dietary habits, assist in planning investment, etc. Rich models have been developed along this line. However, the core of ECA work is not the use of human language technology but the use of facial expressions and gestures in order to bring emotions and politeness in conversation. Research efforts are being made focusing on in-

---

**Figure 1. Example dialogue applications with increasing order of complexity**

<table>
<thead>
<tr>
<th>Complexity</th>
<th>Example Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Long distance dialing</td>
</tr>
<tr>
<td></td>
<td>Train/Bus schedule enquiry</td>
</tr>
<tr>
<td></td>
<td>Travel Booking Agent</td>
</tr>
<tr>
<td></td>
<td>Fixing an electronic circuit</td>
</tr>
<tr>
<td></td>
<td>Auto tutoring</td>
</tr>
<tr>
<td>High</td>
<td>Talking with a person</td>
</tr>
</tbody>
</table>
Related Content

Defect Detection Approach Based on Combination of Histogram Segmentation and Probabilistic Estimation Technique
[www.igi-global.com/article/defect-detection-approach-based-combination/64183?camid=4v1a](www.igi-global.com/article/defect-detection-approach-based-combination/64183?camid=4v1a)

Removal of Pectoral Muscle Region in Digital Mammograms using Binary Thresholding
[www.igi-global.com/article/removal-pectoral-muscle-region-digital/74798?camid=4v1a](www.igi-global.com/article/removal-pectoral-muscle-region-digital/74798?camid=4v1a)

Visual Survey Analysis in Marketing
[www.igi-global.com/chapter/visual-survey-analysis-marketing/48395?camid=4v1a](www.igi-global.com/chapter/visual-survey-analysis-marketing/48395?camid=4v1a)

A Semi-Supervised Metric Learning for Content-Based Image Retrieval
[www.igi-global.com/article/semi-supervised-metric-learning-content/59878?camid=4v1a](www.igi-global.com/article/semi-supervised-metric-learning-content/59878?camid=4v1a)