

# A Modal Logic Framework for Human-Computer Spoken Interaction

L. Villaseñor-Pineda<sup>1</sup>, M. Montes-y-Gómez<sup>1</sup> and J. Caelen<sup>2</sup>

<sup>1</sup> Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Mexico.  
{villasen,mmontesg}@inaoep.mx

<sup>2</sup> Laboratoire CLIPS-IMAG, Université Joseph Fourier, France.  
jean.caelen@imag.fr

**Abstract.** One major goal of human computer interfaces is to simplify the communication task. Traditionally, users have been restricted to the language of computers for this task. With the emerging of the graphical and multimodal interfaces the effort required for working with a computer is decreasing. However, the problem of communication is still present, and users continue caring about the communication task when they deal with a computer. Our work focuses on improving the communication between the human and the computer. This paper presents the foundations of a multimodal dialog model based on a modal logic, which integrates the speech and the action under the same framework.

**Keywords:** human computer spoken interaction, speech acts, multimodal interaction, and modal logic.

## 1 Introduction

The first dialog systems used the speech as the unique communication channel. However, the human communication is strongly *multimodal*. The lips movement, the facial expressions, and the gestures are all of them key elements in the human interchange of information.

Current multimodal dialog systems attempt to integrate several communication modalities along with the speech. The construction of this kind of systems is a complex task [8, 10, 13]. It considers several problems such as: speech recognition, natural language understanding, knowledge representation, fusion of the different input modalities in a coherent message, the definition of a dialog model and others.

This paper focuses on the definition of a dialog model. It presents the foundations of a *multimodal dialog model* based on a modal logic, which represent the rules of the conversation and integrates in the same framework the direct actions (those accomplished with a device of direct designation such as the mouse) and the spoken ones (those orally requested by the user to the machine). This consideration is of great relevance because the spoken actions are not performed immediately such as the direct ones. Thus, the evolution of a spoken action must be controlled during the dialog: from the moment it is proposed until the time it is satisfied.

The proposed model is supported by the theory of speech acts [1, 12]. It is based on the hypothesis that the dialog is conducted by the mental states that maintain the beliefs, desires and intentions of the user. Nevertheless, this model does not attempt to set a human behavior to the machine, but only to give it the logical elements to hold its actions [2]. The idea of treating the speech as an action is not original (see for