Advancing HRI with BILA: A Comprehensive Study

Dai-Long Ngo-Hoang

1VNUHCM

December 7, 2023

Abstract

This paper investigates the pivotal role of artificial languages, with a specific focus on BILA, in addressing challenges in Human-Robot Interaction (HRI). By delving into the complexities of speech recognition and technological constraints, the study introduces BILA as a groundbreaking solution. The research encompasses the design, implementation, and evaluation phases, emphasizing its transformative potential in enhancing user-robot interactions.

Keywords: Human-Robot Interaction, Speech Recognition, Artificial Languages, BILA, Dialogue Management

Introduction

The integration of robots into society mandates a comprehensive exploration of social robotics (Fong, Nourbakhsh, & Dautenhahn, 2003). As robots become integral across various domains, optimizing the user experience takes precedence. In response, researchers in Human-Robot Interaction (HRI) are scrutinizing the intricacies of speech interaction. This paper critically examines speech recognition intricacies, technological constraints, and introduces artificial languages, with a specific emphasis on the innovative BILA (formerly ROILA).

1. Speech in HCI

1.1 Intuitive Modality in Multimodal HCI

Speech Interaction, an intuitive and user-friendly modality within Multimodal Human-Computer Interaction, demands minimal learning efforts. Despite its advantages, challenges arise during errors or breakdowns, significantly impacting the user experience. The balance between user-friendly interactions and system robustness is crucial (Atal, 1995; Slneiderman, 2000; Chen, 2006).

1.2 Complexity of Speech Recognition

Speech recognition involves the conversion of analog sound waves into a digital format, relying on intricate grammar and acoustic models. Challenges include recognizing natural language due to homophones and difficulties in determining word boundaries. These complexities contribute to limitations within current speech recognition technology (Lee, Soong, & Paliwal, 1996).
1.2.1 Overcoming Speech Recognition Challenges

Limitations in current speech recognition technology hinder the widespread acceptance of Speech Interfaces. Issues like homophones and inaccuracies in recognizing word boundaries impact system efficacy, necessitating alternative approaches (Chen, 2006).

1.3 Speech in HRI

1.3.1 Challenges in Dialogue Mapping

In the pursuit of emulating social dialogues within Human-Robot Interaction, challenges in dialogue management and mapping emerge. Irregularities in natural dialogue, complexities in turn-taking, and technological limitations, including synchronization issues, contribute to miscommunication and user frustration (Fry et al., 1998; Churcher, Atwell, & Souter, 1997).

1.3.2 Addressing Technological Limitations

Synchronization issues between the robot’s hardware platform and the speech recognition engine necessitate prioritizing speech as an output modality. Mismatched user expectations and robot abilities result in suboptimal communication, highlighting the need for enhanced technological solutions (Kulyukin, 2006).

1.4 Research Goal

Recent attempts to advance automatic speech recognition technology fall short, emphasizing the need to explore alternative approaches. This paper advocates for a balanced trade-off between user convenience and machine recognition accuracy within speech interfaces (Shneiderman, 2000).

Including the formulas for the BILA robot experiment:

**Semantic Accuracy (SemAc):**
\[
\text{SemAc} = \frac{\text{Number of Correct Sentences}}{\text{Number of Commands}}
\]

**Sentence Accuracy (SenAc):**
\[
\text{SenAc} = \frac{\text{Number of Correct Sentences}}{\text{Number of Commands}}
\]

**Word Accuracy (WoAC):**
\[
\text{WoAC} = 1 - \text{WER}
\]

**Word Error Rate (WER):**
\[
\text{WER} = S + D + 1N
\]

Here, \(S\) represents the number of substitutions, \(D\) is the number of deletions, \(I\) is the number of insertions, and \(N\) is the number of words in the reference sentence.

It’s important to note that the WER formula describes the number of operations needed to convert the recognized sentence into the reference sentence, and \(S\), \(D\), and \(I\) have the same cost.

Additionally, to estimate the time saved when using BILA instead of Vietnamese, you can use the formula:
\[
\text{Time Saved} = 3.5 \times \frac{\text{Word Accuracy}}{100}
\]

For example, if Word Accuracy improves by an additional 18.9%, the time saved would be 3.5 x 18.9 seconds, equivalent to approximately 66.2 seconds (about 1.1 minutes) when speaking BILA instead of Vietnamese.

1.5 Artificial Languages

Artificial languages, designed for communication, play a crucial role in addressing speech recognition challenges. Introducing the innovative BILA (formerly ROILA), this study aims to enhance recognition accuracy. BILA strives to strike a harmonious balance between ease of human learning and machine recognition, offering a fresh perspective on speech recognition challenges (Zimmerman, Forlizzi, & Evenson, 2007).

1.6 Paper Outline

This paper follows the conventional Human-Computer Interaction (HCI) design approach, encompassing investigation, design, implementation, and evaluation. The study critically examines existing artificial languages, details BILA’s design and iterative processes, outlines its implementation in prototypes, and provides insights into user impressions. A comprehensive evaluation in a local school concludes the thesis, summarizing contributions and presenting future prospects for BILA in human-robot interaction.
References: