Title	音声対話処理における社会的要因の重要性
Author(s)	LIU, YUNING
Citation	
Issue Date	2024-12
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/19681
Rights	
Description	Supervisor: 鵜木 祐史, 先端科学技術研究科, 博士

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

Abstract

In daily conversations, participants' behaviors and expressions are influenced by social factors such as roles and status. According to theories like communicative action and communication accommodation, individuals adjust their language and behavior to align with their partner's communication style. This adjustment fosters mutual understanding and enhances the success of the interaction. This natural tendency is both a cognitive process and a crucial social mechanism. Speech entrainment reflects a speaker's desire for social integration or identification with others. Speakers can modulate their entrainment to either reduce social distance by aligning with their partner or increase it through dis-entrainment.

Given the pervasive nature of entrainment in dialogues, this paper's core argument is that entrainment in conversations can reflect certain aspects of social distance. This social distance, in turn, is influenced by the topic and function of the conversation. Previous studies have also indicated that the topic and function in dialogues are part of social factors. Therefore, by quantifying entrainment metrics in conversations, we propose that it is possible to capture and reflect aspects of the social factors present in the interaction. These social factors can deepen our understanding of human interaction and provide objective insights. The main scientific question is how to leverage computational techniques to quantify social factors through entrainment. Unlike previous studies, this research, based on the Interactive Alignment Model (IAM), adopts a top-down theoretical framework and employs a linear model to measure entrainment across multiple acoustic parameters. This approach, compared to traditional methods that calculate entrainment for individual features, better captures sociological factors in conversations. Specifically, this study focuses on exploring how participants employ different strategies to interact across varied scenarios. To test these hypotheses, three research questions are posed: (1) How can various social factors be quantified, and do these quantified factors exhibit different patterns across conversational scenarios? (2) Can these factors help classify or predict conversational scenarios? (3) Is the proposed method of measuring entrainment across multiple parameters superior to traditional methods that consider each feature separately?

To answer these questions, a corpus of Chinese conversations involving scenarios such as arguments, comforting, sharing happiness, and persuasion was developed. Additionally, the open-source IEMOCAP English corpus, containing emotionally rich conversations, was used to validate the findings.

The results demonstrate that speakers adjust their entrainment and dis-entrainment strategies depending on the conversational scenario. The proposed method bridges the gap between psychology and engineering, offering an objective framework for understanding human interaction, thus addressing the first research question. Introducing social factors into both corpora showed that scenario classification in the Chinese corpus achieved 62.3% accuracy, while predicting entrainment trajectories

using acoustic features in the English corpus reached 49.0% accuracy. Integrating acoustic features with social factors improved accuracy by 5% and 2% in the respective corpora, answering the second research question. Finally, comparing the proposed method with traditional entrainment measures validated the third research question, demonstrating the superiority of the combined approach. This paper explores the role of entrainment in human interaction and presents a computational framework to quantify sociological factors. These findings offer new insights into human-to-human interactions and are particularly promising for improving human-machine dialogue systems. By incorporating social factors, future human-machine systems could better perceive and respond to users' social cues, resulting in more natural and engaging interactions.